verbraucherzentrale
Bundesverband

RECOMMENDATIONS TO IMPROVE THE AI ACT

Recommendations of the Federation of German Consumer Organisations (Verbraucherzentrale Bundesverband – vzbv) on the AI Act trilogue negotiations

07.07.2023

Impressum

Bundesverband der Verbraucherzentralen und Verbraucherverbände - Verbraucherzentrale Bundesverband e.V.

Team Digitales und Medien
Digitales@vzbv.de

Rudi-Dutschke-Straße 17
10969 Berlin

Der Verbraucherzentrale Bundesverband e.V. ist im Deutschen Lobbyregister registriert. Sie erreichen den entsprechenden Eintrag hier.

INHALT

I. CORE RECOMMENDATIONS

1. DEFINITION OF 'AI SYSTEM'

- vzbv recommends adopting the European Parliament's proposal for a definition of AI. It aligns with the OECD definition of AI. It is technology neutral, thus future proof and does not run the risk of being too narrow as new technological approaches to AI emerge.

2. NEW CONSUMER RIGHTS AND EFFECTIVE CONSUMER PROTECTION

- vzbv supports the introduction of new consumer rights under the AI Act as proposed by the European Parliament:
  - An obligation for deployers to **inform consumers** when they are subject to the **use of a high-risk AI** system (Article 29 (6) a) (new))
  - The **right to an explanation** when a high-risk AI system produces legal effects (Article 68 c) (new))
  - The **right to lodge a complaint** with a supervisory authority (Article 68a) (new))
  - The right to a **judicial remedy** against a national supervisory authority (Article 68b (new)).
- vzbv supports introducing a **fundamental rights impact assessment for high-risk AI** systems, **involving representatives of affected groups**, including consumer- and civil society organizations (proposal by the European Parliament: Article 68 b (new)).

3. EFFECTIVE ENFORCMENT

- **Inclusion** of the **AI Act to** Annex I of the **European Directive on representative actions for the protection of the collective interests of consumers** ((EU)2020/1828) (proposal by the European Parliament: Article 68d (new)).
- Introduction of a **mandate** for competent **authorities** to **investigate** AI systems if they have **adverse effects** on the **protection of consumers** (proposal by the European Parliament: Article 65 (1)).

4. PROHIBITED PRACTICES

- vzbv welcomes the Council's and the European Parliament's proposals for closing loopholes in the bans of **manipulation** and **deception** (Article 5 (1) a)) and the **exploitation of** people's **vulnerabilities** (Article 5 (1) b) by eliminating the 'intentionality requirement'. vzbv welcomes to include all kinds of harm, not only physical or psychological harm as proposed by the European Parliament. The protection from **exploitation** of persons' **vulnerabilities** must include vulnerabilities due to **age, disabilities**, a person's **characteristics**, **personality** traits and the **social** or **economic situation** as proposed by the Council.
- vzbv welcomes the following additions and specifications regarding prohibited practises:
  - Ban of unjustified AI-based **social scoring** of natural persons **by public** and **private actors** (proposal by the Council and the European Parliament: Article 5 (1) c)).

- Ban of 'real-time' **remote biometric identification** by **public** and **private actors** (proposal by the European Parliament: Article 5 (1) d) (new))
- Ban of AI-based **biometric categorisation** of persons **along sensitive or protected attributes** (proposal by the European Parliament: Article 5 (1) ba) (new)).
- Ban of AI systems creating **facial recognition databases** through scraping of **facial images** from the **internet** or **surveillance cameras** (CCTV) **footage** (proposal by the European Parliament: Article 5 (1) db) (new)).

5. HIGH RISK AI SYSTEMS

- vzbv recommends lawmakers to **adopt** the **original proposal** from the **European Commission** regarding the classification of AI systems as high risk (Article 6). The areas listed in Annex III are already specific use cases. Additional layers with risk-self assessment by providers are not necessary, but introduce legal uncertainty and open the scope and incentive for providers to underestimate the risks.
- The **list of high-risk use cases in Annex III** should be **extended to include** the following use cases:
  - 'Real-time' and 'post' **remote biometric identification** system (proposal by the European Parliament and the Council: Annex III (1)).
  - **Biometrics-based systems** used to make inferences about **personal characteristics** of natural persons based on biometric or biometrics-based data, including **emotion recognition** (proposal by the European Parliament: Annex III (1)).
  - For AI systems **evaluating** the **creditworthiness** of natural persons there should be **no exemption** for AI **systems** used to **identify fraud** (proposal by the European Parliament**:** Annex III (5) b)).
  - AI systems influencing **risk assessment and pricing** in case of **life and health insurance** (proposal by the Council: Annex III (5) ba)).
  - **Content recommender systems** for user-generated content on very large online platforms as defined by the Digital Services Act (proposal by the European Parliament: Annex III (8) ab) (new)).

6. GENERAL PURPOSE AI SYSTEMS, FOUNDATION MODELS, GENERATIVE AI SYSTEMS

- Providers/deployers of **general purpose AI** systems must **comply** with **high risk-AI obligations if** they modify a general purpose AI system such that it **becomes high-risk** (proposal by the European Parliament: Article 28 (1) ba) (new)).
- vzbv supports the introduction of **obligations for providers of foundation models** including **risk assessments** and **mitigation** of **risks involving independent experts**, **data governance** and **system requirements** such as adequate levels of performance, predictability, interpretability, **corrigibility**, safety and cybersecurity (proposal by the European Parliament: Article 28b).
- vzbv supports the **introduction** of due diligence **obligations**, for **providers of foundation models used in generative AI systems** (proposal by the European Parliament: Article 28b (4)).

II. BACKGROUND AND INTRODUCTORY REMARKS

In April 2021, the European Commission proposed a regulation laying down harmonised rules on artificial intelligence (AI) in the (Artificial Intelligence Act (AI Act))[1]. The Council[2] adopted its approach in December 2022. The European Parliament[3] adopted a negotiating position on the AI Act in Brussels in June 2023. The Trilogue has already started. Based on the original proposal of the European Commission, the positions of the Council and the Parliament vzbv has prepared recommendations for the trilogue negotiations.

To accomplish the European Commission's stated objective of a 'trusted AI', vzbv calls for the AI Act to grant consumers strong rights with respect to AI and to strengthen the possibilities for independent assessments of AI systems with the potential to negatively affect consumers, including foundation models. Consumer-friendly regulation of AI has thus become an urgent matter.

III. DETAILED EXPLANATION OF VZBV'S CORE RECOMMENDATIONS

1. DEFINITION OF AI SYSTEM (ARTICLE 3 (1))

The definition of AI should be technology neutral and future proof. Policymakers should therefore avoid an AI definition referring to specific techniques such as "machine learning and/or logic- and knowledge based approaches", as proposed by the Council.

> vzbv recommends adopting the European Parliament's proposal for a definition of AI. It aligns with the OECD definition of AI. It is technology neutral, thus future proof and does not run the risk of being under-inclusive as new technological approaches to AI emerge.

2. NEW CONSUMERS RIGHTS AND EFFECTIVE CONSUMER PROTECTION

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

Consumers need to be informed when a high-risk AI systems is involved in a decision that significantly affects them. Only if they are aware of a high-risk AI involvement, they

---

[1] European Commission: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final) (hereafter 'AIA') (2021), URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206 [Access: 20.07.2021].

[2] Council of the European Union: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach (2022), URL: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf.

[3] European Parliament: Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) (2023), URL: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html [Access: 06.07.2023].

can demand an explanation from the deployer (Article 68 c) (new) AIA)) and duly contest this decision.

> vzbv welcomes the proposal of the European Parliament to introduce a new obligation for deployers to inform consumers whenever they are subject to the use of a high-risk AI system.

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

The consumer's right to be informed that a high-risk system is being used (Article 29 (6) a) is only one side of the coin. Consumers can only exercise their rights and contest AI-based decisions when they receive, upon request, a meaningful explanation from the deployer. That explanation must include information regarding the role of the AI system in the decision-making procedure, the main parameters of the decision taken and the related input data.

> vzbv supports adopting the European Parliament's[4] proposal to introduce a right to explanation of individual decision-making in the AI Act when a high-risk AI system produces legal effects or significantly adversely affects the natural person's health, safety, fundamental rights, socio-economic well-being.

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

It is essential to ensure that consumers have access to justice if AI-associated risks materialise. The right to lodge a complaint with a supervisory authority also provides strong incentives for providers and deployers of AI systems to comply with the AI Acts provisions, especially with the prohibitions laid down in Article 5.

> vzbv supports the European Parliament´s[5] proposal to introduce a right for natural persons to lodge a complaint with a national supervisory authority in the AI Act.

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

Competent authorities have a number of obligations under the AI Act. If they fail to comply with those obligations, particularly the enforcement obligations, the application of the AI Act is seriously threatened (compare for example, how the lack of enforcement of the GDPR affects the application of its rules and principles).

> vzbv supports adopting the European Parliament's[6] proposal to introduce a right for natural persons to an effective judicial remedy against a national supervisory authority in the AI Act.

2.5 Fundamental rights impact assessment for high-risk AI systems (Article 29a (new))

Requiring the deployer of a high-risk AI system to carry out a fundamental rights impact assessment and mitigate the risks is justified. The expenditure for the deployer for carrying out the risk assessment and mitigation is relatively low as compared to the poten-

[4] Amendment 630

[5] Amendment 628

[6] Amendment 629

tial harm a high-risk AI systems can inflicted on natural persons or society. The proposal rightly underlines the importance of contextual and specific uses of AI systems. This will foster consumers' trust in the technology.

> vzbv welcomes the European Parliament's[7] introduction of an obligation for deployers of high-risk AI systems to carry out a fundamental rights impact assessment (Article 29 a (new)).

> vzbv supports the European Parliament's proposal[8] to involve representatives of groups affected by a high-risk AI system in the fundamental risk assessment (Article 29 (a) 4) (new)). However, this involvement must not, as proposed, be limited to public consumer protection agencies but also include civil society organisations representing consumers and other civil society organisations.

3. EFFECTIVE ENFORCEMENT

3.1 Inclusion of the AI Act in Representative Actions Directive (2020/1828/EC) (Article 68d (new))

For effective enforcement, qualified entities must also be able to collectively enforce consumer rights in the case of non-compliance with the AI Act. Individual consumers often cannot enforce their rights vis-a-vis enterprises due to prohibitively high costs and lack of expertise with lengthy court procedures.

> vzbv supports adopting the European Parliament's[9] proposal to include the AIA in the Representative Actions Directive (2020/1828/EC), so that qualified entities can enforce consumer rights in the case of non-compliance with the AI Act.

3.2 Violation of consumer rights: investigation by competent authorities (Article 65 (1))

For adequate consumer protection, competent authorities require an explicit mandate to investigate suspected violations of consumer rights by AI systems. The European Parliament introduces a wider notion of risk in Article 65, including adverse effects on the protection of consumers. This wider notion of risk mandates competent authority's investigations into AI systems, in suspected cases of consumer rights violations.

> vzbv welcomes the introduction of a wider definition of risk by the European Parliament in Article 65 (1)[10]. It allows competent authorities to investigate AI systems if they present a risk for adverse effects on the protection of consumers, the environment, public security, or democracy or the rule of law and other public interests, that are protected by the applicable Union harmonisation law.

4. PROHIBITED PRACTICES (ARTICLE 5)

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

The AI Act should prohibit manipulative or deceptive techniques that harm consumers by impairing a person's ability to make an informed decision.

---

[7] Amendment 413

[8] Amendment 413

[9] Amendment 631

[10] Amendment 596

The Council proposal limits prohibitions to cases where the operator *intentionally* employs *subliminal* techniques *"in order"* to materially distort a person's behaviour, which likely leads to *physical* or *psychological harm* to *that person or another person*. This notion unfortunately provides various loopholes, which the Council's and the European Parliament's proposal aim to close.

(1) Subliminal techniques

Manipulation and deceptive techniques are one of the core risks of consumer-facing AI system in commercial applications, like recommender systems, assistants or marketing tools, aiming at inducing consumers to by products or services. The term 'subliminal technique' is vague and not necessary. In addition, the 'subliminal' criterion implies that the prohibition only applies to techniques, which people can barely notice. This means that Article 5 does not protect consumers from AI-based manipulations that can just be noticed.

> vzbv recommends to delete the wording 'subliminal techniques', as it is vague and not necessary and provides loopholes and to adopt the wording in the European Parliament´s proposal[11] and refer only to "manipulative or deceptive techniques" (Article 5 (1) a)).

(2) Intentionality requirement

vzbv welcomes that the European Parliament and the Council propose to complement the European Commission's intentionality requirement ("in order to") with the qualifier "or the effect of". The European Commission's proposal meant that affected persons had to prove that an operator intentionally harms affected persons, which is near to impossible in practice.

However, the European Parliament reintroduces the intentionally threshold with respect to manipulative or deceptive techniques as it prohibits "subliminal techniques beyond a person's consciousness or *purposefully* manipulative or deceptive techniques").

> vzbv welcomes the proposal by the Council and the European Parliament[12] to complement the intentionality on the side of the operator ("in order to") with the qualifier "or the effect of" in order to eliminate the intentionality requirement. In practice affected consumers or authorities will hardly ever be able to prove that an operator uses manipulative or deceptive techniques *intentionally* to harm consumers ((Article 5 (1) a)). However, vzbv recommends deleting the term "purposefully" in the European Parliaments proposal.

(3) Decision that would not have been taken

> vzbv recommends deleting the term "thereby causing the person to take a decision they would not have taken otherwise" in the European Parliament´s proposal for Article 5 (1) a)[13] as it introduces legal uncertainty and requires a contra factual prove, which is impossible to deliver in practice.

[11] Amendment 215

[12] Amendment 215

[13] Amendment 215

(4) Concept of harm

The European Commission's and the Council's proposal limit the prohibition to *physical* or *psychological* harm. Consumer harm caused by manipulate and deceptive AI techniques is much broader. Especially economic/financial harm should be included.

> vzbv supports the wider definition of harm as proposed by the European Parliament[14]: The words "*physical or psychological"* should be removed from the definition of harm in order to prohibit all kinds of immaterial and material harm resulting from AI-driven manipulations and deception.
>
> vzbv recommends to adopt the proposal of the European Parliament[15] to not only ban the use of manipulative and deceptive techniques distorting the behaviour of *individuals* ("a person") but also for "a group of persons".

4.2 Exploiting peoples vulnerabilities (Article 5 (1) b) (new))

The European Commission's proposal for the ban of exploiting people's vulnerabilities provides various loopholes that the Council and the European Parliament´s proposals address adequately:

The European Commission's ban only applies if the operator of the AI system *intentionally exploits* peoples vulnerabilities of "due to their *age, physical or mental disability*" "*in order to* materially distort their behaviour in a manner that causes that person or another person *physical or psychological harm;*"

(1) Intentionality requirement

> Similarly to the previous article, vzbv welcomes the proposal by the European Parliament[16] and the Council to introduce the words "or the effect of" in order to eliminate the intentionality requirement. In practice, affected consumers or authorities will hardly ever be able to prove that an operator uses manipulative or deceptive techniques to *intentionally* to harm consumers (Article 5 (1) b) (new)).

(2) Vulnerabilities

> vzbv recommends to adopt the European Parliament´s[17] and the Council proposal to ban not only exploitation of a person's vulnerability due to their age, physical or mental disability, but also due to a person's characteristics, known or predicted personality traits or social or economic situation (Article 5 (1) b) (new)).

(3) Concept of harm

> vzbv supports the wider definition of harm as proposed by the European Parliament[18]: The qualifiers *physical or psychological* should be removed from the definition of harm in order to prohibit all kinds of immaterial and material harm that result from AI-based exploitation of personal or group-specific vulnerabilities (Article 5 (1) b) (new)).

[14] Amendment 215

[15] Amendment 215

[16] Amendment 216

[17] Amendment 216

[18] Amendment 216

> vzbv recommends to adopt the proposal of the European Parliament[19] to ban not only the exploitation of a specific group of persons' vulnerabilities, but also of individual persons vulnerabilities.

4.3 Social scoring (Article 5 (1) c) (new))

The intolerable risks of social scoring as defined in the Article 5 (1) c) are not limited to AI systems used by public authorities but also materialise when used by private actors. They include potential exclusion of consumers from important services, discrimination and unfair treatment.

> vzbv recommends to adopt the proposal by the Council and the European Parliament[20] to prohibit unjustified AI-based social scoring of natural persons via AI systems by public as well as private actors (Article 5 (1) b) (new)).

4.4 Remote biometric identification (Article 5 (1) d))

> vzbv recommends to adopt the proposal of the European Parliament[21] to prohibit the use of AI systems for remote biometric identification in publicly accessible spaces for the purpose of law enforcement as well as for private actors (Article 5 (1) d)).

> vzbv welcomes the proposal of the European Parliament to allow an exemption for the use of AI systems for 'post' remote biometric identification of pre-recorded footage when subject to a pre-judicial authorisation for the purpose of law enforcement in cases of serous criminal offenses (Article 5 (1) dd) (new)).

4.5 Categorisation along sensitive or protected attributes (Article 5 (1) ba) (new))

AI systems for biometric categorisation are highly error-prone and if they use sensitive or protected attributes, they are especially problematic and potentially discriminatory.

> vzbv recommends to adopt the proposal of the European Parliament[22] to prohibit AI-based biometric categorisation of natural persons along sensitive or protected attributes or characteristics, with the exception of AI systems intended to be used for approved therapeutic purposes. (Article 5 (1) ba) (new))

This will still allow the use of AI systems for biometric categorisation along *non*-sensitive or *non*-protected attributes, although the AI Act would consider them as high-risk (compare Annex III (1) (new)).

4.6 Facial recognition databases (Article 5 (1) db) (new))

Facial recognition databases using untargeted scraping of facial images from the internet or CCTV footage pose an intolerable risk to all peoples' privacy and personal security, to especially women, e.g. when they are victims of stalkers. Services such as PimEyes[23] scrape the internet for facial images and allow users to upload an image of a

---

[19] Amendment 216

[20] Amendment 218

[21] Amendment 220

[22] Amendment 217

[23] https://pimeyes.com/en

person and search for all other images of that person. This happens without consent of the affected persons.

The same holds for images captured by surveillance cameras (CCTV), many of which are publicly available[24]. The AI Act should ban AI systems allowing strangers to create intimate profiles of movements and online/social media activity of affected persons outright.

> vzbv welcomes the proposal of the European Parliament[25] to ban AI systems that create facial recognition databases through untargeted scraping of facial images from the internet or CCTV footage (Article 5 (1) db) (new)).

vzbv wants to emphasise that limiting the ban to facial images from the internet or CCTV footage might not be sufficient to protect potential victims. Especially the ban of CCTV scraping should include all person-related images. AI systems can identify natural persons from CCTV footage without the need for facial images, only by analysing posture, height, and the personal characteristic way people walk.[26]

5. HIGH-RISK AI SYSTEMS

5.1 Classification rules for high-risk AI systems (Article 6)

The European Commission's proposal for the classification has the advantage of clarity. Unfortunately, the Council and the European Parliament introduce loopholes and legal uncertainty in the classification rules for AI systems.

The Council proposal exempts AI systems from being classified as high risk if their output is 'purely accessory' in respect of the relevant action or decision to be taken. The European Parliament proposes that providers of AI systems can exempt themselves from being classified as high-risk if they consider that their AI system does not pose a significant risk.

The Council and the European Parliament should retain the original proposal by the European Commission. There is simply no need for an additional risk self-assessment by providers to determine which AI system in Annex III constitutes a high risk. The areas of applications listed in Annex III are already relatively narrow, specific use cases. Their potential for causing significant harm is evident. Furthermore, an additional risk self-assessment introduces legal uncertainty and offers large leverage for litigation in courts. This benefits large corporate AI providers to the detriment of SMEs. Allowing providers to self-assess their AI system's risk, introduces a large loophole. Providers have an overwhelming incentive to over-emphasise benefits and downplay risks. As a result, the AI Act will become a toothless tiger.

> vzbv recommends not to introduce an additional risk assessment for classifying AI systems as high risk, but adopt the European Commission's original proposal for Article 6.

---

[24] Insecam.org, URL: http://www.insecam.org/en/ [Access: 11.07.2023].

[25] Amendment 225

[26] Businessinsider.com: China says it has new surveillance camera technology that can recognise you just from how you walk (2018), URL: https://www.businessinsider.com/china-says-new-surveillance-tech-can-id-people-from-their-walk-2018-11 [Access: 11.07.2023].

5.2 Annex III: List of critical high-risk AI areas

vzbv welcomes that the European Parliament and the Council propose to supplement Annex III by additional critical use cases.

(1) Biometric and biometrics-based systems (Annex III (1))

> vzbv welcomes the proposal by the European Parliament[27] and the Council to include 'real-time' and 'post' remote biometric Identification as a high-risk in Annex III (excluding those remote biometric identification systems prohibited under Article 5).

The European Parliament also proposes to include biometrics-based systems as high risk, if they make inferences about personal characteristics of natural persons based on biometric or biometrics-based data, including emotion recognition systems. Such biometric-based systems are highly invasive, error-prone and potentially biased.

> vzbv recommends adopting the proposal of the European Parliament[28] to include AI systems intended to make inferences about personal characteristics of natural persons on the basis of biometric or biometrics-based data as a high-risk area in Annex III (Annex III (1)).

(2) Evaluation of creditworthiness (Annex III (5) b))

AI systems for evaluating a person's creditworthiness present a high risk for affected individuals and should be included in Annex III. At least some AI systems used to identify fraud have proven to be error-prone and discriminatory and caused harm and despair to thousands of persons[29]. The AI Act should consider them as high-risk.

> vzbv supports the proposal by the Council and the European Commission to include AI systems for evaluating a person's creditworthiness in Annex III as both do not contain an exception for AI systems used for the purpose of detecting financial fraud as proposed by the European Parliament[30] (Annex III (5) b)).

(3) Health and life insurance (Annex III (5) ba))

The European Commission's proposal does not include AI systems used in the area of life and health insurance in Annex III at all. The European Parliament[31] proposes to include AI systems influencing decisions on the *eligibility* of natural persons for health and life insurance.

> vzbv welcomes the Council's proposal to consider AI systems as high-risk when they influence the *risk assessment and pricing* in the case of life and health insurance (Annex III (5) b)).

---

[27] Amendment 711

[28] Amendment 712

[29] Netzpolitik.org: Childcare benefits scandal: Dutch government to pay million Euro fine over racist data discrimination (2022), URL: https://netzpolitik.org/2022/childcare-benefits-scandal-dutch-government-to-pay-million-euro-fine-over-racist-data-discrimination/ [Access: 11.07.2023].

[30] Amendment 722

[31] Amendment 723

(4) Recommender systems on very large social media platforms (ANNEX III (8) ab) (new))

Recommender systems used by social media platforms have proved to be a risk to democratic processes and the public discourse as they have a propensity to propagate false information, hate speech and conspiracies.

> vzbv supports adopting the European Parliament's proposal[32] to designate AI systems intended as to be used as recommender systems for user-generated content on very large online platforms as high-risk in (Annex III (8) ab) (new)).

6. OBLIGATIONS FOR PROVIDERS OF GENERAL PURPOSE AI, FOUNDATION MODELS AND GENERATIVE AI

6.1 General Purpose AI

The Council and the European Parliament propose that providers of general purpose AI systems (GPAI), that are used as high-risk systems, shall comply with the requirements applicable to high-risk AI systems in Title III, AIA. The Council's approach offers a large loophole by allowing providers to preclude the use of the GPAI in high-risk areas in the instructions to use "in good faith" (Article 4c (1)). It is impossible that a provider can truly make this claim in good faith.

> vzbv recommends to adopt the European Parliament's proposal[33] that operators of general purpose AI systems must comply with the obligations for providers/deployers of high risk-AI (laid down in Article 16, AIA) if they modify an GPAI such that it becomes high-risk (Article 28 (1) ba) (new)).

6.2 Foundation models

(1) Definition of foundation model (Article 3 (1) 1c) (new))

The European Parliament's proposes to define a foundation model as "an AI system model that is trained on *broad data at scale*". This definition is not future-proof. The focus on data narrows down the definition and ignores other factors such as the number of model parameters. There is a risk, that future foundation models evade this data-focused definition, for example when they rely on less but specific higher quality data[34].

> vzbv recommends to adopt the definition of foundation models as proposed by the European Parliament[35]. But in order to make the definition future proof and technology neutral vzbv recommends to eliminate the term 'broad data at scale'. This is an unclear term, introducing legal uncertainty.

[32] Amendment 740

[33] Amendment 394 introducing Article 28 (1) ba) (new)

[34] The Economist: The bigger-is-better approach to AI is running out of road, URL: https://www.economist.com/science-and-technology/2023/06/21/the-bigger-is-better-approach-to-ai-is-running-out-of-road [Access: 11.07.2023].

[35] Amendment 168

(2) Obligations for providers of foundation models (Article 28b (1), (2), (3) (new))

Some risks from the use of foundation models emerge down the value chain where a deployer applies a foundation model in a specific context. The deployer is best placed to addresses these context-specific risks.[36]

Other risks inherent to foundation models are more general. The developer is best placed (and often the only one) to effectively address these. The AI Act should oblige the original developer/provider to address these risks. If a foundation model is flawed or biased, AI systems down the value chain will "inherit" the flaws of the foundation model.

The European Parliament introduces the obligation for providers of foundation models to identify and mitigate *reasonably foreseeable risks* (Art, 28b (2) a) (new)). This is adequate, as the term *reasonably foreseeable risks* means that providers do *not* have to identify and mitigate *all possible risks* in all imaginable applications, but only the most obvious and relevant ones. For example, precluding that large language models like ChatGPT provide plans for bomb making or encourage suicide.

> vzbv supports the European Parliament's introduction of obligations for providers of foundation models (Article 28b). vzbv welcomes that the obligations include risk assessments and mitigation involving independent experts, data governance and ensuring an adequate data base and system requirements such as adequate levels of performance, predictability, interpretability, corrigibility, safety and cybersecurity.

(3) Specific obligations for providers of generative AI (Article 28 b (4) (new))

Foundation models used as generative AI systems that generate new data, such as text, images, and sound[37], can cause real harm. This includes risks for manipulation and emotional dependence of natural persons, risks to safety, security vulnerabilities and fraud, deepfakes and disinformation, discrimination, privacy and data protection[38].

> vzbv welcomes that the European Parliament introduces specific due diligence and transparency obligations for providers foundation models that are used in generative AI systems (Article 28b (4)).

(4) Involvement of independent experts in risk assessment

The European Parliament introduces the obligation for providers to identify, reduce and mitigate reasonably foreseeable risks "with appropriate methods such as with the involvement of independent experts".

> vzbv recommends to adopt the proposal by the European Parliament to involve independent experts in the identification and mitigation of risks of foundation models.
>
> European authorities defining benchmarks for foundation models according to Article 28b (2) g) (new) and Art 58 a (new) must specify criteria to define the independence

---

[36] If applicable, the deployer must address these by a fundamental rights fundamental rights impact assessment for high-risk AI systems (proposal by the European Parliament: Article 68 b (new) and the obligations for providers of high-risk in Title III.

[37] Prominent examples of text generators are ChatGPT (OpenAI) or Bard (Google), Image generators include for example are Midjourney, Stable Diffusion or DALL-E, Popular example of an audio generator is resemble.ai

[38] For a comprehensive overview of the risks and harms caused by generative AI systems se the recent report by the Norwegian Consumer Council Forbrukerradet: Ghost in the machine – Addressing the consumer harms of generative AI (2023), URL: https://www.forbrukerradet.no/side/new-report-generative-ai-threatens-consumer-rights/ [Access: 11.07.2023].

> experts to ensure that they are truly independent without industry affiliation or vested interests.