Council of the European Union

Brussels, 24 July 2024
(OR. en)

12517/24

Interinstitutional File: 2022/0303(COD)

LIMITE

JUSTCIV 133
JAI 1239
CONSOM 258
COMPET 821
MI 721
FREMP 329
CODEC 1694
TELECOM 249
CYBER 232
DATAPROTECT 264

NOTE

From:	Commission Services
To:	Delegations
No. Cion doc.:	13079/22 + ADD 1+ ADD 2 + ADD 3 + ADD 4
Subject:	Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) - Commission Services considerations on the possible substantive amendments to the AILD in view of the AI Act

Delegations will find in Annex considerations from the Commission Services of possible substantive amendments to the AILD in view of the AI Act.

ANNEX

AI Act – AILD – Commission services considerations on possible substantive amendments

The proposal for an AI Liability Directive (AILD) contains several links with the AI Act (Regulation (EU) 2024/1698). As the AI Act has been agreed with amendments compared to the Commission's proposal, there are several new provisions that could impact the provisions of the AILD.

The approach in the AILD proposal in relation to the AI Act references has been the following:

- To apply the disclosure provisions of Article 3 only to high-risk AI systems, as defined by the AI Act;
- To expressly refer to specific requirements from Chapter III AI Act for providers and deployers of **high-risk AI systems** in paragraphs 2 and 3 of Article 4 AILD only when those requirements are an illustration of the first condition set by paragraph 1 (i.e. when the respective requirement is '**directly intended to protect against the damage that occurred'**).

At the same time, Article 4 paragraphs (1), (5), (6) and (7) AILD apply to all other AI systems than those that are subject to requirements laid down in Chapter III of the AI Act.

The sections below explain the newly added Articles or amended provisions in the AI Act which could be relevant for AILD. They do not constitute a Commission position as regards whether and to what extent the AILD needs to be amended.

I. Article 4 AI Act - AI literacy

The AI Act introduces the concept of AI literacy, which is defined in Art. 3 (56) as the skills, knowledge and understanding needed to make informed decisions about the deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause.

Article 4 AI Act requires providers and deployers of AI systems to take appropriate measures to ensure a sufficient level of AI literacy of their staff and other persons dealing with the operation and the use of AI systems on their behalf. Article 4 AI Act refers to AI systems in general, not specifically to high-risk AI systems. According to Article 113 point (a) AI Act, this obligation will apply six months following the date of entry into force of the AI Act.

The issue arises on whether this obligation is directly intended to protect against the damage a potential claimant may suffer and whether it should be expressly referred to in Article 4 (2) and (3) AILD.

II. Obligations of providers and deployers of high-risk AI systems

The final version of the AI Act introduces three further obligations as regards high-risk AI systems:

a. Article 20 (2) AI Act – investigation of causes of risks

Article 20 (2) AI Act establishes a new obligation for providers becoming aware of a risk within the meaning of Article 79 (1) AI Act, i.e. a risk to health, safety or to fundamental rights of persons. The providers should immediately investigate the causes in collaboration with the reporting deployer, where applicable. They should then inform the market surveillance authorities of the Member States in which they made the high-risk AI system available on the market and, where applicable, the notified body that issued a certificate for that high-risk AI system in accordance with Article 44 AI Act, in particular of the nature of the non-compliance and of any relevant corrective action taken.

The issue arises on whether this obligation is directly intended to protect against the damage a potential claimant may suffer and whether it should be expressly referred to in Article 4 (2) AILD.

b. Article 26 (2) AI Act – human oversight

Article 26 (2) AI Act provides for an obligation of deployers of high-risk AI systems to assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support.

Human oversight is explained in Article 14 AI Act as a feature of the AI system that should be put in place from the design and development phase. The aim of human oversight is to 'prevent and minimise the risks to health, safety and fundamental rights' (Article 14 (2)). While some human oversight measures can be identified and built into the high-risk AI system, others should be implemented by the deployer.

The obligation established in Article 26 (2) AI Act considers the nature of high-risk AI systems and the risks to safety and fundamental rights possibly associated with their use.

For example, if the person who is assigned to oversee the high-risk AI system is trained properly, it is more likely that this person will know when to intervene on the operation of the AI system (see Article 14 (4) (e) AI Act) and if they have the authority to suspend the system they could prevent harm from happening or mitigate it.

The issue arises on whether this obligation is directly intended to protect against the damage a potential claimant may suffer and whether it should be expressly referred to in Article 4 (3) AILD.

c. Article 27 (1) AI Act – Fundamental rights impact assessment for high-risk AI systems

Article 27 (1) AI Act provides for an obligation for some categories of deployers to perform an assessment of the impact on fundamental rights that the use of the system may produce. This obligation only applies to:

- bodies governed by public law or private entities providing public services when using high-risk AI systems referred to in Annex III[1] and
- deployers of high-risk AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score (with the exception of AI systems used for the purpose of detecting financial fraud) and for risk assessment and pricing in relation to natural persons in the case of life and health insurance.

This new obligation has been included during the negotiations of the AI Act. With this obligation, the AI Act recognises that risks can not only stem from the way an AI system is designed by the provider, but also from the way it is used by the deployer. Deployers can therefore best identify potential significant risks that were not foreseen in the development phase (Recital 93 AI Act). Once the impact assessment has been performed, the AI Act provides for an obligation for the deployer to notify the market surveillance authority of the results of the assessment (Article 27 (3) AI Act) by filling a template.

The issue arises on whether this obligation is directly intended to protect against the damage a potential claimant may suffer and whether it should be expressly referred to in Article 4 (3) AILD.

III. General-purpose AI models

General-purpose AI (GPAI) models are defined in Article 3 (63) AI Act as AI models which display significant generality and are capable of competently performing a wide range of distinct tasks and which can be integrated into a variety of downstream systems or applications.

a. Nature of GPAI models and scope of the AI Act in this respect (Recital 97 AI Act):

Recital 97 AI Act explains the nature of GPAI models. The notion of GPAI models must be strictly distinguished from the notion of AI systems. GPAI models are characterised by their generality and the capability to competently perform a wide range of distinct tasks. They are typically trained on large amounts of data and may be placed on the market in various ways, including through libraries, application programming interfaces, as direct download, or as physical copy. GPAI models may be further modified or fine-tuned into new models. They are essential components of AI systems, but they do not constitute AI systems on their own. They require the addition of further components, such as, for example, a user interface, to become AI systems. GPAI models are typically integrated into and form part of AI systems.

[1] With the exception of high-risk AI systems used in areas listed in point 2 of Annex III (i.e. those used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity.)

The AI Act provides specific rules for providers of GPAI models and for GPAI models that pose systemic risks, which should apply also when these models are integrated or form part of an AI system. The obligations for the providers of GPAI models shall apply once these models are placed on the market. When the provider of a GPAI model integrates an own model into its own AI system that is made available on the market or put into service, that model should be considered as being placed on the market and, therefore, the obligations in the AI Act for providers of such models should continue to apply in addition to those for AI systems. The obligations foreseen for providers of GPAI models should in any case not apply when an own model is used for purely internal processes that are not essential for providing a product or a service to third parties and the rights of natural persons are not affected.

When a GPAI model is integrated into or forms part of an AI system, this system should be considered to be a GPAI system when, due to this integration, this system has the capability to serve a variety of purposes. A relevant example for a GPAI system based on a GPAI model is ChatGPT, which is a chatbot and virtual assistant with multi-purpose capabilities.

A GPAI system can be used directly, or it may be integrated into other AI systems (Recital 100). However, the AI Act does not provide for obligations that would apply to providers of GPAI systems specifically on account of the AI system qualifying as a GPAI system. From a regulatory perspective, they are treated like other AI systems. For example, Article 50 AI Act ("Transparency obligations for providers and deployers of certain AI systems") could apply, notably Article 50 (2) dealing with AI systems generating synthetic audio, image, video or text content, which even explicitly refers to GPAI systems. Furthermore, Article 25 (1) point c) addresses situations in which the intended purposes of a GPAI system is modified in such a way that the GPAI system becomes a high-risk AI system.

Finally, it is worth noting the different supervision and enforcement models foreseen for GPAI models and AI systems, including GPAI systems. For AI systems, Article 74 provides for the applicability of the Market Surveillance Regulation, with the result that the supervision and enforcement take place on national level. For GPAI models, the supervision and enforcement are under exclusive competence of the Commission according to Article 88. Given that AI systems frequently integrate GPAI models, the AI Act provides for cooperation between Commission and national authorities in these cases in Article 75 AI Act ("Mutual assistance, market surveillance and control of GPAI systems").

b. Obligations for providers of GPAI models

Article 53 (1) AI Act introduces obligations for providers of GPAI models, such as:

- Drawing up and keeping up to date the technical documentation of the model, including its training and testing process and the results of its evaluation;
- Drawing up, keeping up to date and making available information and documentation to providers of AI systems who intend to integrate the GPAI model into their AI system;
- Putting in place a policy to respect Union copyright law;
- Drawing up and making publicly available a sufficiently detailed summary about the content used for training of the GPAI model.

These obligations primarily aim at providing a comprehensive documentation of the development process, as well as sufficient information to market surveillance authorities and to the public as well as to providers of AI systems that would integrate the GPAI model into an AI system.

In this scenario, GPAI models are essential components of AI systems. AI systems that integrate GPAI models are treated like any other AI systems, meaning that whether they are considered high-risk AI systems follows the rules for the high-risk classification set in Article 6 AI Act. If they are high-risk according to Article 6, the high-risk requirements and obligations of the AI Act fully apply. The intention behind the transparency obligations for providers of GPAI models is to facilitate compliance with the high-risk requirements for providers of high-risk AI systems that use GPAI models as a component.

A relevant example for a high-risk AI system that integrates a GPAI model would be a search tool for recruiters allowing end-user handling through an AI-based chatbot. Such an AI-based chatbot would typically be based on a GPAI model. As the AI system is intended to be used for the recruitment or selection of natural persons, it would classify as high-risk according to Article 6(2) and Annex III, point 4 a). Accordingly, the provider would have to ensure that the system complies with the high-risk requirements of the AI Act. To that end, the provider of the AI system would be entitled to receive information and documentation from the provider of the GPAI model that is integrated into the AI system, for example regarding the data used for training, testing and validation, which may be relevant to meet the data governance requirement of Article 10.

The issue arises on whether these obligations are directly intended to protect against the damage a potential claimant may suffer and whether they should be expressly referred to in Article 4 AILD.

c. Obligations for providers of GPAI models with systemic risk

According to Article 51 (1) AI Act, GPAI models are classified as GPAI models with systemic risk if they have high impact capabilities. Article 3 (65) AI Act defines systemic risk as a risk specific to the high-impact capabilities of GPAI models having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.

According to Recital 110, systemic risks include, but are not limited to, any actual or reasonably foreseeable negative effects in relation to major accidents, disruptions of critical sectors and serious consequences to public health and safety; any actual or reasonably foreseeable negative effects on democratic processes, public and economic security; the dissemination of illegal, false, or discriminatory content. Systemic risks should be understood to increase with model capabilities and model reach, can arise along the entire lifecycle of the model, and are influenced by conditions of misuse, model reliability, model fairness and model security, the degree of autonomy of the model, its access to tools, novel or combined modalities, release and distribution strategies, the potential to remove guardrails and other factors.

Article 51(2) introduces a presumption that GPAI models have such high impact capabilities if the cumulative amount of computation used in the training reaches a certain threshold. The GPAI model classifies *ipso iure* as having systemic risk and providers are obliged to notify the Commission pursuant to Article 52(1). Furthermore, Article 52(4) empowers the Commission to designate GPAI models as presenting systemic risks based on criteria set out in Annex XIII (e.g. number of parameters of the model, quality and size of the data set, or impact due to reach).

Article 55 AI Act introduces further obligations for providers of GPAI models with systemic risk, such as:

- Performing a model evaluation, including conducting and documenting adversarial testing of the model with a view to identifying and mitigating systemic risk;
- Assessing and mitigating possible systemic risks at Union level that may stem from the development, the placing on the market, or the use of GPAI models with systemic risk;
- Keeping track of, documenting and reporting to the AI Office and, as appropriate, to national competent authorities relevant information about serious incidents and possible corrective measures to address them;
- Ensuring an adequate level of cybersecurity protection for the GPAI model with systemic risk and the physical infrastructure of the model.

The obligations introduced by Articles 53 and 55 AI Act serve a public interest by enhancing transparency, accountability and legal compliance. Detailed technical documentation and summaries of training content help stakeholders understand AI development processes, fostering trust. Comprehensive information ensures proper usage and reduces risks. Respecting Union copyright law prevents intellectual property infringements. Performing model evaluations and adversarial testing identifies and mitigates systemic risks before integration of the GPAI model into a variety of AI systems, preventing proliferation of risks. Tracking and reporting serious incidents allows for timely interventions, maintaining public safety. Ensuring robust cybersecurity protection guards against cyber-attacks, preserving the reliability and safety of AI applications. These measures collectively ensure responsible AI development and deployment, safeguarding societal interests rather than addressing individual harm.

As mentioned, above AI systems that integrate GPAI models are treated like other AI systems, meaning that whether they are considered high-risk AI systems follows the rules for the high-risk classification set in Article 6 AI Act. This also applies when the GPAI model is classified as presenting systemic risk.

The issue arises on whether these obligations are directly intended to protect against the damage a potential claimant may suffer and whether they should be expressly referred to in Article 4 AILD.

IV. Article 86 AI Act - Right to explanation of individual decision-making

Article 86 AI Act provides for a right of any affected person to request from the deployer clear and meaningful explanations on the role of the AI system in the decision-making procedure and the main elements of a decision to which that person has been subject.

This right to explanation concerns only decisions taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III[2] (e.g. AI systems used for the use cases in the area of biometrics, education and vocational training, employment, workers management and access to self-employment, access to and enjoyment of essential private services and essential public services and benefits, law enforcement, migration asylum and border control management, administration of justice and democratic processes, with the exception of critical infrastructure).

Moreover, this right can be exercised only if the respective decision produces legal effects or similarly significantly affects the affected person in a way that they consider to adversely impact their health, safety and fundamental rights.
Situations which could fall under Article 86 could include a decision taken in determining access or admission to educational and vocational training or a decision regarding the recruitment or selection of candidates for a job.

The explanations under Article 86 AI Act only concern '*the role of the AI system in the decision-making procedure and the main elements of the decision'*, and not details on how the AI came to produce its output (e.g. what inputs it was given about the affected person or how those inputs are processed in the algorithm so that the specific output about that person was delivered). The explanation received following a request based on Article 86 AI Act could be helpful for the affected person to understand the extent to which the AI output was relevant for the deployer's decision and how it was taken into consideration, thus revealing the internal decision-making processes of the deployer.

The issue arises on whether this right to explanation is relevant for the AILD provisions, in particular Article 3.

---

[2] With the exception of high-risk AI systems used in areas listed in point 2 of Annex III (i.e. those used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity.)