Council of the European Union
General Secretariat

Brussels, 10 October 2022

WK 13628/2022 INIT

LIMITE

TELECOM

WORKING PAPER

This is a paper intended for a specific community of recipients. Handling and further distribution are under the sole responsibility of community members.

CONTRIBUTION

From:	General Secretariat of the Council
To:	Working Party on Telecommunications and Information Society
Subject:	Artificial Intelligence Act - Slovenia non-paper: AI definition

Delegations will find in the Annex Slovenia non-paper: AI definition.

Slovenia non-paper on AI system definition, AI HR classification criteria and assurance of effective remedy

Date: 10. 10. 2022

Clear definition of AI system and AI High-Risk classification rules are crucial parts of AI Act since they present the main requirement that all other parts of the Act rely on. Also, it can be expected that in case of any open issues with respect to the applicability of the AI Act after its adoption in the real world cases, definition and classification will always be the main underlaying questions that will need to be clearly answered. **Slovenia has pointed out the need for clear definition already from the EU White paper on AI in 2018. With this respect, the main goals that we think we need to strive for are:**

- clear definition of AI system in order to bring clarity foremost on technical level to understand what is regulated, for all relevant stakeholders;
- interoperability of AIA for wide international reach;
- clear definition of scope of AIA, where we do not think the right approach is to do this through definition of AI system itself, but instead we argue to do this by clearly describe properties/context of use of AI systems that we want to regulate;
- proper structure of AIA so that it can be adapted based on quick developments in the AI area;
- clear approach towards not regulating technology, but rather its use.

Based on current draft proposal, Slovenia argues that additional effort needs to be put to clearly define AI, criteria and context of AI system that we want to regulate with the scope of AIA and criteria according to which those systems we want to regulate are regarded as high risk AI (HR AI). We think that we currently have tools available to successfully and effectively do this. We argue for:

1. using OECD AI system definition because we think it is technically sound and understandable and enables international compatibility and reach of AIA, which requires using **OECD's AI system's constitutive parts, i.e. learning, modelling, reasoning in definition of AI system in the main part of AIA** (article 3),
2. using **OECD AI classification framework to define HR AI** in order to have much more **granular capability distinguish high risk AI use cases** from other use in order to adapt to real world scenarios (article 7),
3. In addition, we would like to raise the need to **assure that proper information is at hand to all stakeholders and flows through value chain** (from providers to users) in order for users to provide all the information to the persons affected by use of AI system, so that they can exercise their **effective remedy procedure** in case they feel their rights have been violated. We acknowledge that liability regime is not part of the AIA but rather separate directive (just adopted by European Commission), but no effective remedy can be assured if affected persons do not get proper information they need, which we need to assure within AIA.

1. AI system definition

Slovenia sees the AI definition as one of the central points of AIA, which we have commented already from the first feedback to EU White paper on AI and the first

feedback on draft AIA. Slovenia has from the beginning argued for the use of the OECD definition of AI system, since we think it is clear and internationally harmonized and accepted definition of AI system, as the basis for the definition of AI system in the AI Act. We propose to describe AI system definition based on the OECD AI system detailed conceptual view (Figure 1) and to describe its main parts as clearly as possible with appropriate text. **We believe that previous text in SI/FR draft proposals in Art 3(1) reflects this conceptual view in a clear way** since it includes **the main defining constitutive properties of AI** system (also from the point of view of used Annex I techniques and approaches) in definition itself in art. 3**, i.e. perception of the world (input), constructing model representing the problem space (learning), refining the model according to the application needs (modelling), inferring information/decisions based on the model used (reasoning) and acting back to the world (output) – all using different techniques and approaches (previously defined in Annex I).** If we understand that any system implemented in SW/HW have input and output functionalities, AI systems' main distinction is that they encompass **learning, modelling and reasoning** in implementation of required solution which **enables execution of a system based on models that have not been pre-defined** (hard-coded SW), so the AI system can give an output also for inputs that have not been programmed in advance.

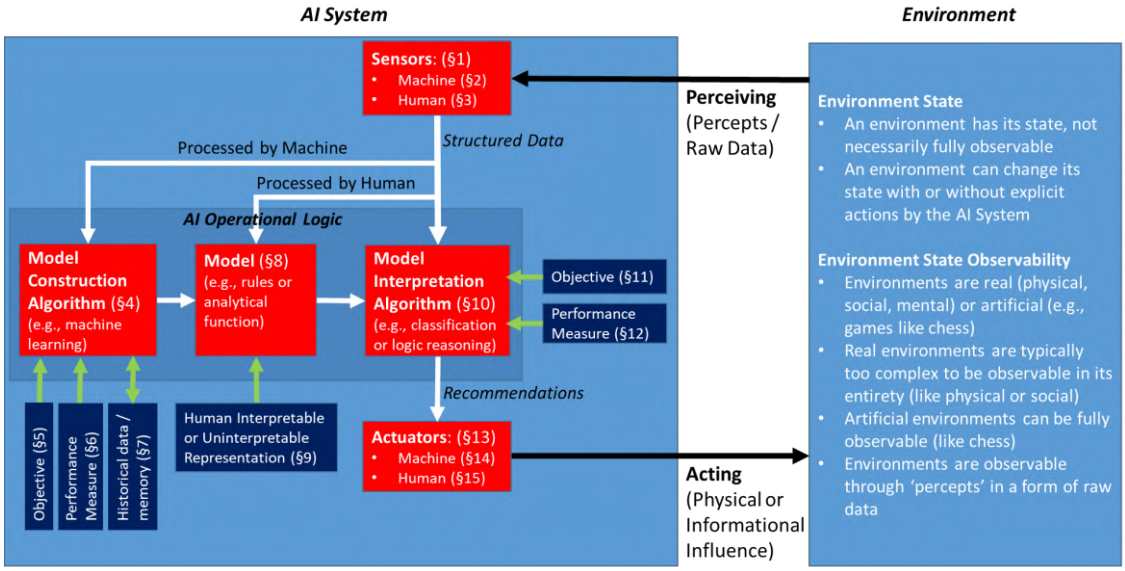

Figure 1 AI system detailed conceptual view (@OECD AIGO)

This is why we think that these three main functionalities, i.e. **learning, modelling and reasoning are the crucial part of the definition of AI system also because they are easily understood to wider audience that AIA is targeting and because the working of many AI solutions can in such a way be easier to explain**. Definition itself need to be foremost technically clear and sound in order to be used to build other necessary frameworks on top (HR requirements framework, risk assessment framework, liability framework, etc.) where understanding of inner working of an AI system together with external interfaces of a system with environment will need to be clear as much as possible.

With respect to recent change of the text of AI definition in art. 3 by CZ draft proposal, removing reference to techniques and approaches defined in Annex I and instead explicitly including some of them, namely “machine learning and knowledge -and rule based approaches” directly in the definition, we understand some of the argumentation for such change but we think that **additional clarifications** about **to what exactly these approaches relate to** need to be clarified. We think that **we need to explicitly reference that these approaches relate to implementation of learning, modelling and reasoning** within any AI system**.** Namely, from the current text proposal it can be **understood that only machine learning techniques (to implement learning) is the defining part of an AI system, so that machine learning alone can be used to describe all sorts of AI systems**. In addition, it can also be **understood that only “logic- and knowledge based approaches” are taken into account when implementing machine learning (or any AI solution)**. Slovenia thinks that **this is not the case**, so we argue to **retain the notion of all three functionalities, i.e. learning, modelling and reasoning in the AI system definition,** so that **it is clear that not only machine learning** (as an approach for implementation of learning) **is defining part of AI system.** In addition, we also argue that **not only “logic- and knowledge based approaches” are those that define AI system**, but rather **many other approaches such as** different **computational approaches**, which can not be described by “logic- and knowledge approaches” (e.g. different sorts of neural networks; stochastic approaches such as Bayes, K-nearest neighbours; statistical approaches such as regression; etc.) **and which can and are widely used today and they will also be used in the future for implementation of AI systems.** What is important is that **all these approaches are not used only to implement specific machine learning method but are typically used in implementation of any of the three main functionalities, i.e. learning (i.e. model construction), modelling (i.e. setting and refining the model of problem space) and reasoning (i.e. model interpretation), many times implemented in an integrated manner or interchangeably**.

With this in mind we do not think that specific approach itself can be used as the defining part of an AI system, rather we think that the main defining part is actually description of **core functionality that AI solution implements which in our view encompass learning, modelling and reasoning (according to OECD model). We think this is better description to clearly define AI system and distinguish it from other systems.**

We do agree that use of machine learning for learning/constructing the model of a problem space is one of the main approaches that are used today widely in AI systems and is one of the important factors that brings non-deterministic nature to current AI solutions we want to regulate, but we think that “system that uses rules defined solely by natural persons to automatically execute operations«, as explained in recital 6, still presents AI system (for example historically old experts systems), even if we do not want to have them regulated. They still can have non-deterministic nature implemented either within modelling or reasoning. We also acknowledge that traditional statistical solutions do not necessarily present AI system, but statistical methods can not be excluded from techniques and approaches to implement AI systems since they are widely used in current AI systems.

Slovenia is not against the goal to limit the scope of which AI systems we would like to regulate in AIA if this is needed or requested by members states. But we think that the policy choice to select the set of AI systems to be regulated (together with the choice of available legislative process for adaptation) **should not be done by redefining current definition of AI system (excluding 3 main constitutive parts) in art. 3, which as such is based on internationally accepted OECD definition, but rather by (1)**

description/limitation of internal properties of AI system, including approaches and techniques used to implement it (if needed) and (2) description/limitation of context in which AI system is used together with the impact that it can have on its environment in specific use case scenarios – namely, impact on safety and health and/or human rights (as proposed to be the main design goal of a risk based framework of the AIA). The later can in principle be included in art. 2 where scope of AIA is defined[1]**.**

With the scope in mind we do acknowledge the need to distinguish AI from more classic SW and programming as explained in CZ proposal in recital 6 which is why we remind ourselves of the reason for regulating AI in contrast to other sorts of SW applications. We agree it is because AI system exhibits properties such as opacity, complexity, autonomy and/or learning/evolving ability based on new data (as explained by Commission in the Impact assessment accompanying AIA) - **so overall non-deterministic behaviour** that can result in **unexpected impact** of using such systems over time and thus **producing undesirable consequences on health, safety and human rights and freedoms**. We very much agree with description in recital 6a that AI system are typically used to **solve complex problems** either where there is no suitable formalisation of the problem solution (e.g. no analytical way to find a solution) or the solution can not be practically found because of for example computational (asymptotic) complexity (it might for example require too much time/space to calculate the solution). This means that with such problems we need to rely on methods that do not give us formally correct solution for specific input, but rather some sort of approximation of it, but which is however still good enough for the targeted purpose. **But those methods do not include or rely on machine learning alone, but modelling and reasoning as well. They are not based only on logic- and knowledge based approaches, but typically also different sorts of other approaches such as computational approaches mentioned above. If we define AI systems only by using machine learning, many AI system based on evolutionary computing approaches (e.g. genetic algorithms), heuristic search and optimization methods, AI planning, etc. would not be covered although they can also exhibit non-deterministic nature that can pose threats to safety, health and human rights and freedoms, which is why we want to regulate AI systems.**

This is why we think the current definition of AI system which is based only on "machine learning and logic- and knowledge based approaches" provides less clarity as to what AI system is as compared with previous definition which references learning, modelling and reasoning as defining part of AI system. We do acknowledge the additional description in recitals 6a and 6b, but feel that this is not enough, since we agree with those member states that requested that the **main definition of AI system need not be described in recitals or annexes, but be clearly defined and included in definition of AI system in the main part of the regulation - articles (art. 3)**.

We also agree with many members states' positions that the definition needs to be structured in such a way that also **reflects the current and future specific techniques used in AI systems** and **provides for the efficient possibilities for its update and adaptation according to the AI developments in the future – at least in the early phase of learning curve at the beginning of AIA implementation**. This is closely **linked to the available process that AIA provides for such adaptation**. More specifically, this is linked to the ability of EU to adapt the AIA according to new AI

[1] Current proposal to limit the scope in AI definition to machine learning only can be related to criterion IV(b) of OECD classification framework, requirements for autonomy can be related to criterion V(c), etc.

development using appropriate process. Slovenia has argued that **structure and process linked to AI definition in art. 3 and ANNEX I and ability to change ANNEX I by delegated act within initial period of AIA implementation of 5 years properly reflected these policy goals**. If Annex I is omitted in such a way that from all techniques and approaches mentioned only machine learning and logic- and knowledge based approaches are taken on board and written directly in the definition, **we feel that we get less clarity with respect to definition of AI system and less capability to adapt to future needs.** Current proposal to specify the technical elements of approaches in the implemented act (art. 4) somehow solves the problem of adaptability only partially, but in our opinion, since it covers **only machine learning and logic- and knowledge based approaches,** still provides for less flexibility as described earlier. From this perspective we think that previous structure of AI system definition (art. 3 + ANNEX I) suits this purpose better than the current CZ proposal omitting the Annex I although we acknowledge that some member states do not support wide use of delegated acts.

The main changes we propose:

4. to explicitly include main defining part of AI system (learning, modelling, reasoning) in the AI system definition and not limiting approaches only to learning/machine learning, but including also computational approaches – both linked to Annex I, also do not support exclusion of human-defined objectives (art. 3);
5. ex an example to include proposal for excluding AI systems where model is specified and hard-coded up-front before AI system is placed on the market or put into service and where such AI system produces only output (on specific input) that has been programmed to produce, so to not bring any non-deterministic behaviour related to specific input during its lifetime use (art. 2);
6. to include both logic- and knowledge based approaches and computational approaches as techniques to implement AI system (and not only learning functionality) either by retaining and possibly narrowing techniques and approaches in Annex I or if Annex I is to be omitted at least by including both logic- and knowledge based approaches and computational approaches as techniques to implement AI system and thus defined in implemented acts (art. 4).

Change suggestion (proposed changes in yellow):

Article 2
Scope

8. This Regulation shall not apply to AI systems where:

7. internal model used to solve a problem that AI system as a whole is set to solve in order to achieve its intended purpose is not generated (learned) from data using machine learning approaches, but specified and hard-coded by human during development of AI system before the AI system is placed on the market or put into service, and

8. no computational approaches are used for model refining (modelling) and model interpretation (reasoning) during the use of AI system, which would result in producing outputs that have not been already pre-defined or programmed at the time of development of the AI system.

Article 3
Definitions

(1) 'artificial intelligence system' (AI system) means a system that is designed to operate with a certain level of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of human-defined objectives using **learning, modelling and reasoning implemented with** ~~machine learning and/or~~ **data-driven approaches,** logic- and knowledge based approaches **and/or computational approaches listed in Annex I**, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts;

Article 4
*~~Amendments to Annex I~~**Implementing acts***

~~The Commission is empowered to adopt delegated acts~~ **In order to ensure uniform conditions for the implementation of this Regulation as regards ~~machine learning approaches and~~ data-driven approaches, logic- and knowledged based approaches and computational approaches referred to in Article 3(1), the Commission may adopt implementing acts to specify the technical elements of those approaches, taking into account market and technological developments. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 74(2).** ~~in accordance with Article 73 to amend the list of techniques and approaches listed in Annex I **within the scope of the definition of an AI system as provided for in Article 3(1)**, in order to update that list to market and technological developments on the basis of characteristics that are similar to the techniques and approaches listed therein.~~

ANNEX I
TECHNIQUES AND APPROACHES
referred to in Article 3, point 1

(a) **Data-driven approaches, focusing on various forms of** Machine learning ~~approaches,~~ including supervised, **semi-supervised,** unsupervised and reinforcement learning, **batch and incremental learning,** using a wide variety of methods including **logic-based learning (such as inductive logic programming)** and deep learning **(discriminative and generative models);**

(b) Logic- and knowledge-based approaches, including knowledge representation **such as ontologies and knowledge bases/knowledge graphs,** ~~inductive (logic) programming, knowledge bases,~~ inference, **abductive** and deductive engines, symbolic **or rule based** reasoning and **decision support/**expert systems;

(c) **Computational approaches including evolutionary computing,** Statistical **and probabilistic** approaches **(including different regression methods and** Bayesian estimation**), different forms of** search and **both single- and multi-objective** optimization methods.

2. Criteria for High Risk AI systems

The Republic of Slovenia advocates for a clear definition of the criteria for determining cases of AI systems that require stricter evaluation and supervision, i.e. high-risk AI and, in this context, as clear as possible the definition of the categories of AI systems actually covered by the AI Act, which must be specifically reflected in the criteria for defining high-risk AI. We believe that current and specially even more complex future landscape of AI technologies and systems require more sophisticated framework to assess whether specific use of an AI system poses high risk (Article 7(2)), thus falls together with providers and users under AI Act's requirements for high-risk systems. In this regard, the

Republic of Slovenia has previously drawn attention to the need to separate requirements and criteria for assessing risk to at least those that relate to AI system inner properties (e.g. requirements on input data, ML models, decision models, model properties such as explainability, bias, transparency, result and response of the AI system, etc.) and requirements related to the environment or context of use in which the AI system operates (e.g. human supervision, record keeping and data keeping, information provisioning, use case scenarios, level of autonomy, impacted persons, etc.).

With this in mind, Slovenia has already proposed to align to and use criteria that can define different AI system use landscape in much greater granularity as it is proposed in Art 7(2) by using **OECD Framework for the Classification of AI systems**[2] (**Figure 2**). This framework classifies specific AI system (based on the main OECD AI system definition - please look above) taking into account **5 main categories** that include both **inner properties of AI system (AI model), different scenarios and capabilities with respect to input (Data & input) and output (task & output) of AI system, different context in which AI system is used including economic context as much as impact and interaction that AI system has on people and environment**. This framework would provide **much more flexibility and precision** in specifying HR AI systems and **thus better adaptation to real world use cases and scenarios when deciding which AI system use cases pose higher or lower risks**. To be more illustrative, when looking at AI system usage landscape we **want to use 4K resolution instead of plain VGA**. Consequently, this would **bring more clarity** for all involved and possibly regulated stakeholders as to what and how specific use of AI system is regulated. The exact wording of classification criteria can be elaborated further if needed to clarify specific points.

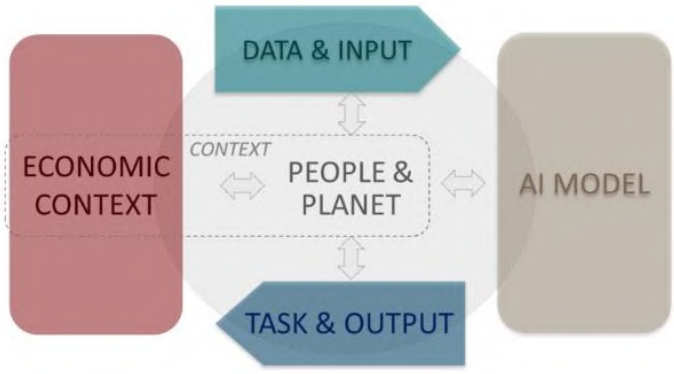

Figure 2 Key high-level dimensions of the OECD Framework for the Classification of AI Systems

The main changes we propose:

9. to extend criteria used when assessing whether an AI system poses a risk of harm to the health and safety or a risk of adverse impact on fundamental rights (art. 7(2)) either by leaving them in art 7(2) or for better clarity and readability adding them all together in new Annex X.

[2] https://www.oecd-ilibrary.org/science-and-technology/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en

Change suggestion (proposed changes in yellow):

Article 7
Amendments to Annex III

1. The Commission is empowered to adopt delegated acts in accordance with Article 73 to ~~update~~ **amend** the list **of high-risk AI systems** in Annex III by adding high-risk AI systems where both of the following conditions are fulfilled:

(a) the AI systems are intended to be used in any of the **application** areas listed in points 1 to 8 of Annex III;

(b) the AI systems pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights, that is, in respect of its severity and probability of occurrence, equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III.

2. When assessing for the purposes of paragraph 1 whether an AI system poses a risk of harm to the health and safety or a risk of adverse impact on fundamental rights that is equivalent to or greater than the risk of harm posed by the high-risk AI systems already referred to in Annex III, the Commission shall take into account the ~~following~~ criteria **specified in Annex X**:

~~(a) the intended purpose of the AI system;~~

~~(b) the extent to which an AI system has been used or is likely to be used;~~

~~(c) the extent to which the use of an AI system has already caused harm to the health and safety or adverse impact on the fundamental rights or has given rise to significant concerns in relation to the materialisation of such harm or adverse impact, as demonstrated by reports or documented allegations submitted to national competent authorities;~~

~~(d) the potential extent of such harm or such adverse impact, in particular in terms of its intensity and its ability to affect a plurality of persons;~~

~~(e) the extent to which potentially harmed or adversely impacted persons are dependent on the outcome produced with an AI system, in particular because for practical or legal reasons it is not reasonably possible to opt-out from that outcome;~~

~~(f) the extent to which potentially harmed or adversely impacted persons are in a vulnerable position in relation to the user of an AI system, in particular due to an imbalance of power, knowledge, economic or social circumstances, or age;~~

~~(g) the extent to which the outcome produced with an AI system is easily reversible, whereby outcomes having an impact on the health or safety of persons shall not be considered as easily reversible;~~

~~(h) the extent to which existing Union legislation provides for:~~

~~(i) effective measures of redress in relation to the risks posed by an AI system, with the exclusion of claims for damages;~~

~~(ii) effective measures to prevent or substantially minimise those risks.~~

3. The Commission is empowered to adopt delegated acts in accordance with Article 73 to amend the list in Annex III by deleting high-risk AI systems where the following conditions are fulfilled:

(a) the high-risk AI system(s) concerned no longer pose any significant risks to fundamental rights, health or safety, taking into account the criteria listed in paragraph 2;

(b) the deletion does not decrease the overall level of protection of health, safety and fundamental rights under Union law.

ANNEX X
CRITERIA FOR ASSESING HIGH RISK AI SYSTEMS
referred to in article 7, point 2

When assessing AI systems for the purposes of Article 7(2), the Commission shall take into account the following criteria:

(i) criteria related to socio-economic context, including its broader natural and physical environment such as sector in which an AI system is deployed, its business function, its critical (or non-critical) nature, its deployment impact and scale:

(a) the intended purpose of the AI system;

(b) the scale to which an AI system has been used or is likely to be used **such as sector specific use, multi sector use, region or country specific use, cross-border and international use;**

(c) industrial sector that has implications in terms of industry structure, regulation, and policymaking for AI systems;

(d) business function and model such as use in recruitment, promotion, training, marketing, procurement, logistics;

(e) capability to impact critical functions, infrastructure or activities of which the interruption or disruption would have serious consequences on;

(f) technology maturity based on TRL representing potential for large scale deployment.

(ii) criteria related to the question of how people and environment as a whole interact with or are affected by an AI system throughout its lifecycle:

(a) required level of AI competency of the user such as amateur, trained practitioner, AI expert;

(b) type of impacted stakeholders such as workers/employees, consumers, business, government agencies or regulators, scientists or researchers;

(c) degree of choice that users or impacted stakeholders have on whether to be subject to the effects of an AI system or not such as whether users can opt out of the effects or the influence of the AI system or challenge, correct or reverse the AI system's output ex-post:

(1) the extent to which potentially harmed or adversely impacted persons are dependent on the outcome produced with an AI system, in particular because for practical or legal reasons it is not reasonably possible to opt-out from that outcome;

(2) the extent to which potentially harmed or adversely impacted persons are in a vulnerable position in relation to the user of an AI system, in particular due to an imbalance of power, knowledge, economic or social circumstances, or age;

(3) the extent to which the outcome produced with an AI system is easily reversible, whereby outcomes having an impact on the health or safety of persons shall not be considered as easily reversible;

(4) the extent to which existing Union legislation provides for:

(i) effective measures of redress in relation to the risks posed by an AI system, with the exclusion of claims for damages;

(ii) effective measures to prevent or substantially minimize those risks.

(d) **type and scale of benefits and risks to human rights and democratic values**

(1) the extent to which the use of an AI system has already brought benefits or caused harm to or adverse impact on the fundamental rights or has given rise to significant concerns in relation to the materialization of such harm or adverse impact, as demonstrated by reports or documented allegations submitted to national competent authorities;

(2) the potential extent of such benefits or harm or such adverse impact, in particular in terms of its intensity and its ability to affect a plurality of persons;

(e) **type and scale of benefits and risks to environment, well-being and society**

(1) the extent to which the use of an AI system has already brought benefits or caused harm to the health and safety or has given rise to significant concerns in relation to the materialization of such harm, as demonstrated by reports or documented allegations submitted to national competent authorities;

(2) the potential extent of such benefits or harm, in particular in terms of its intensity and its ability to affect a plurality of persons;

(f) potential for human labour displacement based on ability of an AI system to automate tasks previously been, or currently conducted by humans.

(iii) criteria related to data and input to AI systems which can be generated by humans and/or automated tools and relate to provenance, the data collection method and origin, their technical characteristics, domain and data quality and appropriateness:

(a) detection and collection method of data and inputs (e.g. by humans or machines);

(b) provenance of data and input (e.g. expert input, provided data, observed data, synthetic data, derived data);

(c) dynamic nature of data (e.g. static, dynamic, real-time);

(d) scale of data (gigabytes, petabytes, exabytes);

(e) intellectual rights associated with data and inputs (e.g. proprietary, personal, public);
(f) identifiability or personal data (e.g. identified, pseudonymized, anonymized);
(g) data quality and appropriateness ensuring that the data are appropriate for use in a project i.e. fit for purpose, and relevant to the system or process following standard practice in the industry sector;
(h) structure and format of data and input (e.g. unstructured, semi-structured, structured).

(iv) criteria related to technical properties of AI system itself based on AI model or multiple models together with techniques and approaches used to implement the AI system for achieving its intended purpose:

(a) AI model characteristics including availability of information about AI model used, rights associated with model (e.g. proprietary, open-source, public), AI model type (e.g. symbolic, statistical, hybrid) and AI model purpose (e.g. discriminative, generative);

(b) AI model building process typically using machine-learning or human-encoding of knowledge – "in the lab", guided by objectives (e.g., output variables) and performance measures (e.g., accuracy, resources for training, and representativeness of the dataset);

(c) AI model evolution capabilities – "in the field", continuing to evolve / acquire abilities from interacting directly with data used in the model building process;

(d) AI model learning type (e.g. centralized or distributed/federated "at the edge");

(e) AI model development and maintenance (e.g. universal single pre-trained model, customizable model that can be re-trained using different data, tailored model);

(f) AI model inference approach – "using a model" to derive output based on "new" data that the model was not trained on, using different inference strategies usually designed to optimize specific objectives and performance measures such as robustness, accuracy, speed, business metrics or other criteria, based on different approaches (e.g. deterministic, probabilistic, combined);

(g) AI model degrees of transparency and explainability.

(v) criteria related to task that AI system performs and output that AI system provides:

(a) task performed by AI system that drives the AI model choice and refers to what the system does, i.e. the function that it performs such as recognition, event detection, forecasting, personalization, interaction support, goal-driven optimization and reasoning;

(b) multi-task, composite AI system capability for combining of tasks and actions into multi-task, composite systems that executes them before producing an output that influences the environment;

(c) system autonomy level and degree of human involvement (e.g. human support, human-in-the-loop, human-on-the-loop, human-out-of-the-loop) to perform an action that influences the environment in which the system operates;

(d) core application areas such as human language technologies, computer vision, robotics, automation or optimization;

(e) evaluation methods and standards available to assess AI system for specific task and context.

3. Ensure proper information flow to enable effective remedy for persons affected by high-risk AI systems

In terms of accountability which is important Ethics Guidelines requirement, we think that regulation should assure **proper information flow from provider to user and user to third party as affected persons** so that user/consumer or affected persons can have all necessary information to report problems with the AI system and obtain a solution, so to ensure the right for effective remedy.

In order to assure proper use of AI system, provider needs to provide proper information in technical documentation (Annex IV) also with respect to the effected persons (we propose to use this term consistently throughout the AIA). User needs to get this information (article 13) both about the **way AI system interacts with environment** and **how to use the AI system safely for himself and people affected by it** (taking into account both intended use as much as **reasonably foreseeable misuse**). It is **crucial that user has certain information that they get from the provider (art 13)**. Specifically foreseeable circumstances and performance and accuracy of AI system that may lead to risks to the health and safety or fundamental rights and freedoms **of him or third persons**. This is **crucial for obligation of a user to also properly inform affected persons about functionating of AI system (art 29)** when third persons' safety, health and fundamental rights and freedoms may have been violated **so they can start investigation or redress of the decisions taken**.

AIA starts from the point that **safety, health and fundamental rights and freedoms need to be protected for all people affected by AI system**, so the proposed change suggestions of art 13, 29 and Annex IV are crucial in order for regulation to enable this goal.

The main changes we propose:

10. more clearly define the goals of transparency provisions in art. 13 with respect to the users of AI system or are affected by it,
11. add clear obligation of users to provide information to affected persons in art. 29,
12. including reference to affected persons in technical documentation in Annex IV.

Change suggestion (proposed changes in yellow):

Article 13
Transparency and provision of information to users

1. High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent ~~to enable users to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured,~~ with a view

to achieving compliance with the relevant obligations of the user and of the provider set out in Chapter 3 of this Title **and enabling users to: ~~understand and use the system appropriately.~~**

(a) understand how the system interacts with the environment where it is intended to be used according to its intended purpose or under conditions of reasonably foreseeable misuse in order to correctly interpret the system's behaviour; and

(b) use the system safely and appropriately in accordance with its intended purpose or under conditions of reasonably foreseeable misuse which may lead to risks to the health and safety or fundamental rights and freedoms.

2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users.

3. The information referred to in paragraph 2 shall specify:

(a) the identity and the contact details of the provider and, where applicable, of its authorised representative;

(b) the characteristics, capabilities and limitations of performance of the high-risk AI system, including:

(i) its intended purpose**, inclusive of the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used**;

(ii) the level of accuracy, **including its metrics,** robustness and cybersecurity referred to in Article 15 against which the high-risk AI system has been tested and validated and which can be expected, and any known and foreseeable circumstances that may have an impact on that expected level of accuracy, robustness and cybersecurity;

(iii) any known or foreseeable circumstance, related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, which may lead to risks to the health and safety or fundamental rights and freedoms of the user or persons or groups of persons on which the system is intended to be used or are affected by the use of the system **referred to in Aricle 9(2)**;

(iv) **when appropriate,** its ~~performance~~ **behaviour regarding specific** ~~as regards the~~ persons or groups of persons on which the system is intended to be used or are affected by the use of the system;

(v) when appropriate, specifications for the input data, or any other relevant information in terms of the training, validation and testing data sets used, taking into account the intended purpose of the AI system~~.~~;

(vi) when appropriate, specifications of the outputs of the system in order to correctly interpret the system's possible impact on the user or persons or groups of persons on which the system is intended to be used or are affected by the use of the system, taking into account the intended purpose of the AI system.

(c) the changes to the high-risk AI system and its performance which have been pre-determined by the provider at the moment of the initial conformity assessment, if any;

(d) the human oversight measures referred to in Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of AI systems by the users;

(e) **the computational and hardware resources needed**, the expected lifetime of the high-risk AI system and any necessary maintenance and care measures to ensure the proper functioning of that AI system, including as regards software updates~~.~~;

(f) a description of the mechanism included within the AI system that allows users to properly collect, store and interpret the logs, where relevant.

Article 29
Obligations of users of high-risk AI systems

6b. Users of high-risk AI systems shall use the information provided under Article 13 to provide sufficient information and explanations to the third persons or groups of persons that AI system is used on or are affected by the use of AI system to safeguard the rights and freedoms of such

third persons or groups of persons in order to allow for the effective protection of their rights where an AI system may have caused harm to their health, safety or fundamental rights and freedoms.

ANNEX IV
TECHNICAL DOCUMENTATION referred to in Article 11(1)

2. A detailed description of the elements of the AI system and of the process for its development, including:

(b) the design specifications of the system, namely the general logic of the AI system and of the algorithms; the key design choices including the rationale and assumptions made, also with regard to persons or groups of persons on which the system is intended to be used or are affected by the use of AI system; the main classification choices; what the system is designed to optimise for and the relevance of the different parameters; the decisions about any possible trade-off made regarding the technical solutions adopted to comply with the requirements set out in Title III, Chapter 2;

3. Detailed information about the monitoring, functioning and control of the AI system, in particular with regard to: its capabilities and limitations in performance, including the degrees of accuracy for specific persons or groups of persons on which the system is intended to be used or are affected by the use of AI system and the overall expected level of accuracy in relation to its intended purpose; the foreseeable unintended outcomes and sources of risks to health and safety, fundamental rights and discrimination in view of the intended purpose of the AI system; the human oversight measures needed in accordance with Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of AI systems by the users; specifications on input data, as appropriate;