Upcoming European Regulations on Artificial Intelligence and Cybersecurity

Markus Dominik Mueck, Amit Elazari Bar On, Stephane Du Boispean
Intel Corp.
{Markus.Dominik.Mueck, Amit.Elazari.Bar.On, Stephane.du.Boispean}@intel.com

***Abstract*—The European Commission is in the process of fundamentally revising the regulatory framework and related market access conditions in key technological areas, including Artificial Intelligence as well as Digital Technology in general. In the present paper, we provide an overview on the status of related policy actions, specifically addressing the novel upcoming Artificial Intelligence (AI) Act and Cyber Resilience Act (CRA) initiatives. Finally, an outlook is given on architectural choices which will help manufacturers to comply with the upcoming new requirements and thus maintain access to the European Single Market.**

Keywords—Artificial Intelligence, Cyber Security

I. INTRODUCTION

Currently, the European Commission (EC) is driving a number of regulatory initiatives which are highly relevant to the industry. Those regulations introduce new essential requirements for given product categories. Conformity to those essential requirements is required for accessing the European Single Market.

As a first key initiative, the European Commission is in the process of finalizing a regulation on Artificial Intelligence (AI) technology. A first draft of the related AI Act is currently available [1]. Among others, the objective of the Regulation is to ensure that fundamental rights of EU citizens are being guaranteed, especially in case AI systems are applied in critical fields ("high risk" AI systems), such as biometric identification, critical infrastructure, etc.

Secondly and in complement to the AI Act, a Cyber Resilience Act (CRA) [2] is currently available in its draft version targeting "*products with digital elements*" including software and hardware products and related components. CRA is expected to serve as a horizontal regulation, likely integrating some of the provisions of the Radio Equipment Directive [3] (Cybersecurity and Privacy related) in the future and also relating to the AI Act as illustrated in Fig. 1.

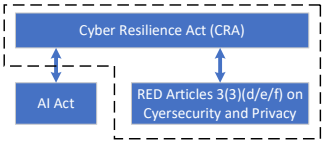

Figure 1: Relationship between CRA, AI Act and RED.

Those regulations are under the New Legislative Framework (NLF), relying on the following implementation steps (simplified):

1) A regulation is issued, typically following a suitable consultation procedure;

2) European Standardisation Organisations (ESOs ETSI (European Telecommunications Standards Institute), CENELEC (European Committee for Electrotechnical Standardization) and CEN (European Committee for Standardization)) are receiving a Standardisation Request (SR) issued by the European Commission.

3) ESOs develop Harmonised Standards (HSs) and possibly other deliverables in support of the regulation, including a definition of technical requirements as well as a test framework for ensuring compliance with the essential requirements of the regulation. After publication of a corresponding reference in the EU Official Journal, compliance with the HSs typically grants presumption of conformity with the regulation and is thus typically the preferred tool used by manufacturers to validate market access requirements. The authors recommend that stakeholders engage in the HS development process in the relevant ESOs (ETSI, CENELEC, CEN) in order to shape the technical details of market access conditions according to industrial needs.

Although the above mentioned regulatory initiatives are limited to the European Single Market, it is expected that related requirements will affect corresponding debates in all regions and will have a major influence on technology regulation world-wide.

The remainder of the paper is organized as follows: Section II will give an overview on the upcoming AI Act [1] followed by a summary of the current status of the Cyber Resilience Act discussions in Section III. Section IV introduces a technical proposal for meeting some of the requirements outlined in the new regulation initiatives, followed by a conclusion in Section V.

II. Overview of the Artificial Intelligence Act (AI Act)

The AI Act is currently available as a draft [1] and is stating the following objectives:

1. ensure that AI systems placed on the European Union market and used are safe and respect existing law on fundamental rights and European Union values;
2. ensure legal certainty to facilitate investment and innovation in AI;
3. enhance governance and effective enforcement of existing law on fundamental rights (e.g., GDPR [5]) and safety requirements applicable to AI systems;
4. facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

Concerning the definition of an AI System, the draft AI Act complements the definitions of the Organisation for Economic Co-operation and Development ("*a software that is developed with one or more of the techniques and approaches listed in Annex I [1] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with*") with at least one of the three main paradigms of 'intelligence' (see Annex I [1]). It is noticeable that this definition is still under debate in the European Parliament and Council and may be generalized to any system instead of focusing on software only.

The draft AI Act [1] introduces a number of Articles – some of those relating to specific technical requirements are summarized in Tab. 1. The requirements need to be met by any AI System which is considered "High Risk" as further defined in the Annex of the AI Act; examples include the *Management and operation of critical infrastructure: AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity*. It is currently under debate in the European Parliament and Council whether cellular communication systems will be considered "High Risk". A corresponding decision may have far reaching consequences for manufacturers of cellular infrastructure and mobile devices. A further open question is how to validate the correct implementation of the newly defined functional requirements by a manufacturer; we cannot rely on classical conformity assessment approaches where physical measures are being verified (such as output power limitations, etc.).

In order to support the work of ESOs, the Joint Research Center of the European Commission (JRC) has analyzed related available standards in an "AI Watch" activity [6]. In particular, standards developed by ISO/IEC as well as ETSI have been identified as being a suitable base-line of the future work. Still, many of the available specifications are of rather generic nature and require further processing in order to be able to offer specific technical requirements and related compliance testing procedures.

It is expected that the AI Act will finally be published in 2023 or early 2024 and related Standardization Requests will be issued to ESOs (i.e., ETSI, CENELEC and CEN) in two stages: i) an initial Standardization Request is expected for early 2023 with the objective to develop supporting deliverables other than HSs; ii) then, as soon as the AI Act is finalized and published, a second Standardization Request is expected to follow, tasking ESOs to develop HSs in support of the AI Act and thus providing a useful tool to industry to validate compliance of products to ensure access to the European Single Market.

Requirements	Summary as defined by the AI Act [1]
Data and data governance	High-risk AI systems … shall be developed on the basis of training, validation and testing data sets that meet the quality criteria ...
Technical documentation	The technical documentation shall be drawn up in such a way to demonstrate that the high-risk AI system complies with the requirements …
Record keeping	High-risk AI systems shall be designed and developed with capabilities enabling the automatic recording of events ('logs') …
Transparency and information to users	High-risk AI systems shall … ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately …
Human oversight	High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use …
Accuracy robustness and cybersecurity	High-risk AI systems shall … achieve, in the light of their intended purpose, an appropriate level of accuracy…
Risk management system	A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems …
Quality management system	Providers of high-risk AI systems shall put a quality management system in place that ensures compliance with this Regulation …

Table 1: Draft AI Act Articles (Selection) related to Technical Requirements [1].

Draft CRA Annex III defines 23 products of class I including Operating systems not covered by class II, routers, modems intended for the connection to the internet, and switches, not covered by class II, Microprocessors not covered by class II, Microcontrollers, application specific integrated circuits (ASIC) and field-programmable gate arrays (FPGA), etc.
Draft CRA Annex III defines 15 products of class II including general purpose microprocessors, microprocessors intended for integration in programmable logic controllers and secure elements, routers, modems intended for the connection to the internet, and switches, intended for industrial use, secure elements, Hardware Security Modules (HSMs), smart meters, etc.

Table 2: Examples of critical products class I and class II as defined by draft CRA [2].

III. Overview of the Cyber Resilience Act (CRA)

The European Commission published a proposal of the Cyber Resilience Act (CRA) [2] in September 2022. It is creating cybersecurity related conformity assessment requirements for so-called "*products with digital elements*" which are defined to be "*any software or hardware product and its remote data processing solutions, including software or hardware components to be placed on the market separately*". Furthermore, the following product categories are being introduced by Annex III of the CRA: i) critical products class I, ii) critical products class II and iii) other (non-critical) products. Corresponding examples are summarized in Tab. 2. Also, a "highly critical product" category is being prepared for possible future usage.

As illustrated in Fig. 1, the CRA is expected to play a horizontal role relating in particular to the AI Act (Article 8 of the Draft CRA is detailing its relationship to High-Risk AI Systems) and integrating part of the Radio Equipment Directive [3]. More specifically, the draft CRA states that it applies "*to all radio equipment within the scope of Commission Delegated Regulation (EU) 2022/30 [4]*" and the CRA requirements "*include all the elements of the essential requirements referred to in Article 3(3), points (d), (e) and (f)*" of the Radio Equipment Directive. In order to avoid regulatory overlap, it is foreseen that the European

Commission will "*repeal or amend the Delegated Regulation [4] with respect to the radio equipment covered by the proposed Regulation, so that the latter one would apply to it, once applicable"*. As of now, the exact implementation is unclear but it is expected that relevant HSs may be transferred from the Radio Equipment Directive to become applicable under the CRA. Essential requirements are defined in Annex I of [2] as they need to be met by manufacturers of products and components in scope to access the European Single Market. They are organized in the following two requirements categories:

1. Security requirements relating to the properties of products with digital elements;
2. Vulnerability handling requirements.

While other regulations such as the Radio Equipment Directive relate to the "placing on the market" of equipment, it is noticeable that CRA addresses i) *placing a product with digital elements on the market*, and ii) *for the expected product lifetime or for a period of five years*. The notion of requirements over the product lifetime will require further discussions, since manufacturers may no longer be in full control of their products after being placed on the market and sold to customers.

It should be noted that the scope of the future Regulation, especially regarding critical products (as defined in Annex III [2]) includes equipment which until now was considered as components rather than products under the NLF harmonized legislations such as the Radio Equipment Directive. Manufacturers of these components are therefore likely to become manufacturers under the EU product safety framework and should prepare accordingly.

Most of the CRA provisions are fully in line with the 2008 NLF decision, which illustrates the intention to ensure regulatory consistency and coherence. Provisions on notifying authorities, notified bodies, obligations of manufacturers and importers are similar to the already existing provisions of the NLF, including the RED. Presumption of conformity is included in the draft Regulation, although some further clarifications on some of the provisions would be useful: A main change proposed by the CRA is the restriction of the conformity assessment procedures. For critical products, the NLF module A (also known as Self-Declaration of Conformity) would not be possible anymore. Only a type-examination performed by a Notified Body, combined with an internal production control (modules B + C) or a full quality assurance (module H) would be allowed. This raises the issue of the conformity of products which have been placed on the market under the RED before the adoption of the CRA. It also requires a swift adoption of HSs by the ESOs, in time for the entry into force of the CRA provisions.

IV. System Architecture supporting AI Act

In order to meet the requirements of the draft AI Act [1], the authors of the present paper propose one possible approach to a novel AI system architecture as illustrated by Fig. 2. This architecture reflects the main functions introduced as summarized by Tab. 1 and further accommodates for user interaction and the connection of the AI system to a database that may be used for the provision of suitable reference training data, logging of user actions, logging of AI system behavior, etc. Note that CRA and RED related requirements are complementary but may be included in the Entity for Risk Mitigation – meeting Cybersecurity and Privacy related requirements indeed rely on a risk-based approach. The various entities of the AI system architecture are summarized below.

A. Entity for AI Processing

The "Entity for AI Processing" is the core of the AI system and is typically being trained using an appropriate training data set and optionally some additional data that is being acquired, while the AI system is being operated. Such an AI system can rely on various machine learning methods / approaches including: regression, classification, clustering, dimensionality reduction, ensemble methods, neural networks and deep learning, transfer learning, reinforcement learning, natural language processing and word embeddings.

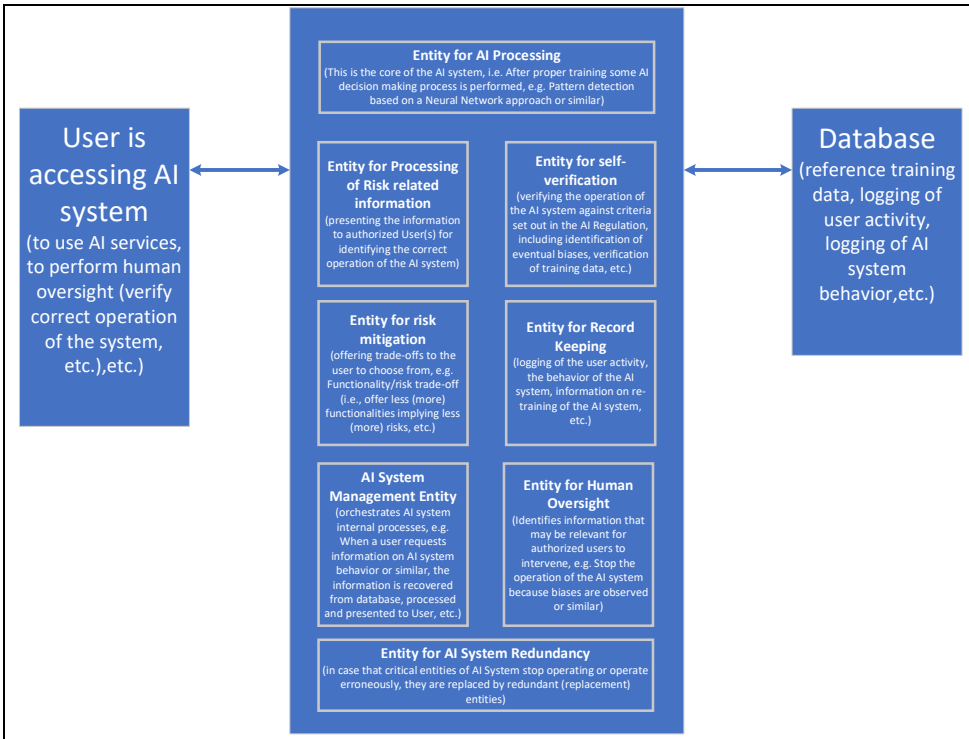

Figure 2: Entities of an AI system in support of the AI Act [1].

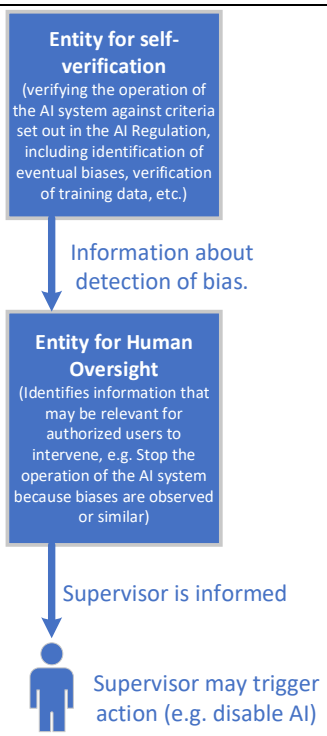

Figure 3: Example of interactions between proposed entities of an AI System (avoidance of biases).

The entity for AI processing is typically using a model that is trained through suitable training data provided by the attached database. Furthermore, the correct operation of the AI system is monitored and controlled through an authorized supervisor. In case that any undesired behavior is being detected, several possible steps may be taken: for example, the supervisor may trigger a retraining of the model using authorized and error-free training data, may report related behavior to the manufacturer, etc.

The entity for AI processing is typically interacting with all other entities of the AI system as further detailed below. This interaction ensures that all requirements of the AI Act are being met – starting with market introduction of the AI system and furthermore including the permanent supervision during the operation of the AI system (e.g., with the objective to identify any introduction of biases into the decision making processes of the AI system).

B. Entity for Self-Verification

The "Entity for Self-Verification" is proposed to verify the correct operation of the newly trained system. It is, in particular, proposed to use a pre-defined test-data set (which is different from the data set used for training) as input to the AI system in order to verify the correct operation. Only if the correct operation is verified, the AI system is allowed for full usage for its intended purpose. In the opposite case, e.g., in case of any unexpected behavior, the operation of the AI system is interrupted, until the issues are resolved. Such a verification step is periodically repeated in case of retraining of the AI system with new data.

A key requirement of the draft AI Act relates to the avoidance of undesired biases, which requires, among others, that diverse user groups are being treated equally. It is indeed possible that any AI system develops such biases during on-going

retraining processes – those need to be detected in the earliest stage possible and suitable counter measures need to be taken. One possibility is to put the system back into a predefined state by applying approved and verified training data to derive the AI model.

C. *Entity for Record Keeping*

When the system is finally used for its intended purpose (after all successful verification steps), the "Entity for Record Keeping" will be logging all user interactions (commands given by the authorized user, etc.) and will record the behavior of the AI system and store relevant information in the database.

D. *Entity for Risk Mitigation*

The "Entity for Risk Mitigation" will propose a trade-off between risk and functionality to the user. For example, the entity may propose that the system is constantly retrained using the observed information obtained during operation. The upside is that this may improve the quality of the AI decision making. The risk is that the new data may introduce biases or other undesired characteristics.

E. *Entity for Processing Risk*

The "Entity for Processing Risk of related Information" will take the results of other entities, such as the "Entity for Self-Verification" and process identified risk related information and unexpected behavior information, such that it can be presented in a concise way to the authorized user

F. *Entity for Human Oversight*

The "Entity for Human Oversight" will allow the authorized user to take action in case that the AI system operates in an unexpected or undesired way, e.g., in case that the decision-making processes indicate undesired biases. The user may then take several actions, including termination of the AI system operation, enforce a retraining of the system, choose a different risk trade-off through the "Entity for Risk Mitigation", etc.

G. *Entity for AI System Redundancy*

The "Entity for AI System Redundancy" oversees redundant replacement options for critical entities of the AI system. In case that some malfunctioning entity is identified, typically relying on information by the "Entity for Self-Verification", then this entity is used to configure a corresponding replacement. After the replacement, the correct operation of the AI system is typically verified, again relying on information by the "Entity for Self-Verification". If it is successfully verified, then the operation of the AI system may continue.

H. *AI System Management Entity*

The "AI System Management Entity" will orchestrate the interaction between the different building blocks indicated above. For example, when one of the entities of an AI system is dysfunctional or operates in an unexpected way, then the "AI System Management Entity" may detect this behavior relying on information by the "Entity for Self-Verification" and may trigger the replacement of concerned entities by redundant replacement entities through the "Entity for AI System Redundancy".

Besides the possibility of interacting with an integrated database, the AI system may be interacting with external (independent) entities operated by 3rd parties.

I. *Interactions between Entities*

The interactions between the various entities introduced above depend on the specific use cases. Fig. 3 provides an example for the case of the detection of undesired biases. The Entity for Self-Verification is responsible for the detection of any undesired biases, will inform the Entity for Human Oversight and finally an authored supervisor will be able to take appropriate action, e.g. terminating the AI operation.

V. Next Steps and Conclusion

The AI Act [1] and the Cyber Resilience Act [2] will decisively impact the access to the European Single Market for technological products – introducing new requirements to be met by industry for continued access to the European Single Market.

For industry, it is essential that a participation to the development of underlying Harmonised Standards continues to be possible through all ESOs as it is foreseen in the New Legislative Framework (NLF). Non-compliance of products would lead to a loss of access to the European Single Market. Also, further clarification is needed on the relationship between the different pieces of legislation. Assuming that there may be a difference in the conformity assessment procedures, industry requires guidance on how to proceed. Finally, we recommend a close linkage to international standards development in support of the EU initiatives, including for development of Harmonised Standards on globally relevant aspects.

Finally, the authors acknowledge that the AI Act as well as the Cyber Resilience Act are currently in a draft stage and may evolve. While we do not expect that the main principles will change substantially, certain refinements are likely to be implemented.

VI. ACKNOWLEDGMENT

This work has been funded by the European Commission through the H2020 project Hexa-X (Grant Agreement no. 101015956).

REFERENCES

[1] 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, April 2021

[2] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020, issued on 15th September 2022

[3] DIRECTIVE 2014/53/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC

[4] Commission Delegated Regulation (EU) 2022/30 of 29 October 2021 supplementing Directive 2014/53/EU of the European Parliament and of the Council with regard to the application of the essential requirements referred to in Article 3(3), points (d), (e) and (f), of that Directive (Text with EEA relevance)

[5] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

[6] AI Watch, AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework, Joint Research Center of the EC, 2021

MARKUS D. MUECK (markus.dominik.mueck@intel.com) received the Dipl.-Ing. and ing. dipl. degrees from the University of Stuttgart, Germany and the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France and the Doctorate degree of ENST in Communications; he is heading the Next Generation and Standards organization of Intel Deutschland GmbH, Munich, Germany and acts as Vice-Chair of the ETSI Board and as Chair of ETSI's OCG-AI committee coordinating Artificial Intelligence activities across ETSI; he is member of the Board of the 5G Automotive Association (5GAA) and Adj. Professor of University of Technology, Sydney, Australia. Dr. Mueck furthermore leads the activities of Europe's 6G Flagship Project Hexa-X on Artificial Intelligence and Machine Learning.

STEPHANE DU BOISPEAN (stephane.du.boispean@intel.com) received his M.A. degrees from the Free University of Berlin and from the Institut d'Etudes Politiques (SciencesPo), Paris, France. He is currently Director, Government Affairs EU & France at Intel, where he leads Intel's engagement towards Governments on industrial policy, smart mobility and product safety.

AMIT ELAZARI BAR ON (amit.elazari.bar.on@intel.com) is a Director of Global Cybersecurity Policy at Intel Corporation and a Lecturer at UC Berkeley's School of Information Master in Information and Cybersecurity. She holds a JSD from UC Berkeley School of Law and graduated summa cum laude three prior degrees. Her research in information security law and policy has appeared in leading technology law journals, presented at conferences and featured at leading news sites. In 2018, she received a Center for Long Term Cybersecurity grant for her work on private ordering regulating information security, exploring safe harbors for security researchers. She practiced law in Israel