Recommendations on the European Commission's WHITE PAPER on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final (the "AI White Paper")

by Vasile Ţiple

Recommendation 1 – Establishing the Principles of AI Regulatory Framework

The EU Artificial Intelligence Principles

The current definitions given by the Commission and the High-Level Expert Group do not differentiate between the different types of AI. Making such a distinction is important, as different types of AI may require different approaches in terms of preparation for their implementation and safeguards which need to be put in place.

The development of a legal and technical regulatory framework for AI in the EU should focus on three principles (the "**AI Principles**")*:*

1. *Understanding* the potential of AI. At this moment we can distinguish three types of AI: Basic AI (which are capable of autonomous data gathering and analysis, task and process automation, document understanding, machine learning, etc.); Autonomous AI (capable of executing tasks while also autonomously learning to become more efficient, fast and accurate and at a more superior level than we currently have under the Basic AI stage) and Advanced AI (which will be able to process and generate independent solutions to the field in which it is deployed and use independently various technologies currently operated by humans to achieve the objectives set by the humans or even efficiency and process improvements as defined by its own learning tools). It is imperative that we make the distinction between them even at this early stage through research and analysis of already existing data and systems. This will give us the right tools to make predictions with respect to the impact of AI, develop and prepare the necessary framework, including from a regulatory perspective, for possible scenarios and new modus operandi in various established activities brought by AI once certain key infrastructure elements for AI are developed (both software and hardware based, i.e. 5G network, quantum computers[1], neuromorphic solutions[2]);
2. *Preparing* for full-scale AI Systems (of all types) being adopted and implemented in Europe, including for the disruption this will cause; and
3. Putting in place the necessary *safeguards* required to maintain the designed AI framework. Such measures will be essential to ensure integrity of EU developed AI Systems when these will be correlated or integrated within a regional or even global AI Systems. The goal is to ensure compliance with any minimal legal and technical

[1] Quantum computers will have the capacity to process in less than seconds many fold larger data sets than today's highest performance computers allowing for the development of new AI applications across sectors.
Source: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[2] Neuromorphic solutions means any very large-scale system of integrated circuits that mimic neuro-biological architectures present in the nervous system.
Source: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

principles and controls established by the EU. This will enable a truly ethical, trustworthy and secure development and adoption of AI in the EU, consistent with the values and rights of EU citizens. We need to ensure that AI Systems designed in the US or Asia, which will be integrated with EU developed systems, or used within the EU use the same core values in their autonomous decision-making process. For example, in case of an AI System for self-driving cars, we may want to ensure that all such cars, regardless of where they are manufactured, allow the driver to override the AI System in similar circumstances.

Recommendation 2 – Structural Framework

Creating a European Artificial Intelligence Agency (EAIA)

Current efforts on developing a legal and technical regulatory framework are coordinated at a high level by the European Commission through the High-Level Expert Group on Artificial Intelligence (AI HLEG) it created which is tasked with drafting of Ethics Guidelines for Artificial Intelligence and AI Policy and Investment Recommendations. The European AI Alliance set up in parallel to the AI HLEG as broad multi-stakeholder forum that also provides input from the different parts of society to the work of the AI HLEG and EU policy-making more generally. This approach seems fragmentary and insufficient for given the potential impact of AI on society.

A more efficient and productive approach would be setting up a dedicated EU Agency (European Artificial Intelligence Agency - EAIA). The EAIA would be in charge with:

- *Research*. The EAIA would have teams dedicated to the research of AI and its likely impact on society, in order to prepare the right pre-emptive regulations and interventions from a technical, legal, financial, ethical/moral levels;
- *Coordination*. The EAIA would be tasked to ensure the coordination within the EU space of the efforts conducted by other EU and national institutions, as well as those of the private sector;
- *Collaboration*. The EAIA would also ensure the collaboration on AI matters with any relevant institutions and the private sector from outside the EU;
- *Budget*. The EAIA could also have a budget dedicated to the support of innovation and initiatives in the AI field through various programs, grants, scholarships etc.
- *Implementation*. The EAIA should also be able to actively participate and follow up on the implementation of its initiatives and objectives.

Based on the envisaged importance of AI and its disruptive capabilities for the society and day to day life, it is vital to allocate the right resources and people for the proper development of this industry.

Five reasons explain why EU agencies are very efficient in implementing EU policies. They (1) tackle problems with a cross-border dimension, (2) pool together EU expertise, (3) provide independent advice to the EU legislator, (4) offer cost-efficient solutions and (5) also increase the Union's visibility by being spread across Member States.[3] We should also add here that

[3] European Policy Brief, No. 52, September 2018, p.2 *More Room for European Agencies in the EU*

an EU Agency would have the capability to tackle issues related to its activity not only with a cross-border dimension but also inter-institutional dimension within the EU itself.

European Artificial Intelligence Agency (EAIA) Local Departments

The EAIA's activity would be carried out through departments specialized on each of the industries which will likely be impacted by the AI, such as: healthcare and pharma, infrastructure, law, economy, energy, environment, etc., with dedicated teams working in each department. Each department would be organized as a Center of Excellence in AI for its specific focus area, such as for example Environment Protection AI Center of Excellence, Energy & Sustainability AI Center of Excellence, etc. Further, each of the EU 44 existing agencies could also have its own team involved in the relevant EAIA department in order to ensure the right synergy and consolidation of all available resources for tackling a specific subject within their area of competence, such as the food industry, medicines, chemicals, education, environment, justice to EU citizens fundamental rights and consumer related industries.

This approach would effectively create the infrastructure for implementing the European Commission's action no. 2 described in its AI White Paper: "*the Commission will facilitate the creation of excellence and testing centres that can combine European, national and private investments, possibly including a new legal instrument. The Commission has proposed an ambitious and dedicated amount to support world reference testing centres in Europe under the Digital Europe Programme and complemented where appropriate by research and innovation actions of Horizon Europe as part of the Multiannual Financial Framework for 2021 to 2027."* This action would need to be completed with the possibility to use these centres of excellence also as the *venue for international collaboration* in AI as they would hold the main EU based know how, expertise and logistics.

The EAIA's sources of funding would be a dedicated budget from the EU, as well as any financial support given by the private sector.

Recommendation 3 – EU AI Objectives

The main objectives of the EAIA would be:

1. Cooperation. Working together with leading companies, start-ups, research labs, etc. to create certain technical and legal standards for the development and implementation of AI solutions. This would be one of the first steps required to make sure the AI Principles are being applied in order to develop AI Systems tailored to the EU's and its Member States' needs, while ensuring that there is a consistent view of the bigger picture among the EU's institutions and across the EU Member States. Creating the right standards around the AI Principles should not be confused with over-regulating the subject matter. The standards should be about creating a flexible, predictable and adaptable framework which could enable and encourage EU innovators, start-ups and companies to develop within the EU framework and prevent

---

Decision-Making Process? Basile Ridard.
Source: http://www.egmontinstitute.be/content/uploads/2018/09/EPB52.pdf

an exodus from the EU to other, potentially, better legal or business frameworks, while at the same time encouraging an influx of external innovators, start-ups and companies to invest in the EU.

2. Centres of Excellence. EAIA will support the EU's digital transformation by creating the Centres of Excellence through which sectorial smart industries[4] initiatives can be researched, developed, implemented and monitored in an organized and trackable manner against the objectives established in the EU's AI and digital strategies.

3. Institutional Understanding. Developing *institutional knowledge and skills* at EU level from the start by having EAIA and its departments pioneer mechanisms to learn new skills and competencies relevant to the new technical developments which (after applying the AI Principles) will cover all the industries impacted by each of the three AI types described at Recommendation 1 above. It is important to distinguish between such AI types in order to ensure that proper human support, logistic and resources are allocated. The EU needs to ensure that the right expertise exists within the EU agencies, its institutions, as well as within the Member States' authorities. This can only be achieved through the right education and on the job training and re-skilling where needed. For the proper execution of this objective a *dedicated* EU agency such as EAIA is required.

4. Evolutionary AI Based Society. We are currently in the presence of *Basic AI Systems* which are rapidly evolving towards *Autonomous AI Systems*, the ultimate destination being *Advanced AI Systems* the capabilities and impact of which we can only suspect at this point by corroborating separate elements of technologies which are now independently in development. It will take a certain amount of time until a more integrated ecosystem will take shape, such as a comprehensive interlinked systems currently called Internet of Things, Autonomous Hardware powered by Autonomous AI Systems and other physical machinery, neuromorphic solutions, quantum computing, evolution and possible impact of cryptocurrencies, implementation of blockchain contracts, the future of computational law, etc. It is still unclear how all these elements will fit together: if some will be excluded, incorporated into some other category or complete or compete with one another. The impact of blockchain on how contracts may be executed, cryptocurrency on how payments are made or if computational law[5] has a future in ensuring mechanical legal compliance

[4] UN Commission on Science and Technology for Development, Nineteenth session, Geneva, 9–13 May 2016
Item 3(a) of the provisional agenda, Smart cities and infrastructure, Report of the Secretary-General. Source: https://unctad.org/meetings/en/SessionalDocuments/ecn162016d2_en.pdf

[5] Computational Law, The Cop in the Backseat, Michael Genesereth, CodeX: The Center for Legal Informatics, Stanford University, Abstract: "Computational Law is that branch of legal informatics concerned with the mechanization of legal analysis (whether done by humans or machines). It emphasizes explicit behavioral constraints and eschews implicit rules of conduct. Importantly, there is a commitment to a level of rigor in specifying laws that is sufficient to support entirely mechanical processing. While the idea of mechanized legal analysis is not new, its prospects are better than ever due to recent technological developments - including progress in Computational Logic, the growth of

independent of human intervention are matters that require at this point specific granular analysis by taking at the same time a holistic approach to ensure the bigger picture is not omitted while dwelling into the weeds of each branch of these and other new technologies. The EAIA would be tasked to understand and contribute to how all the pieces of the puzzle will fit together.

5. Coordinated Adoption. Action 6 of the AI White Paper proposes that the public sector within the EU and Member States adopt AI technologies through sector dialogues aimed at preparing a specific 'Adopt AI programme' which will support public procurement of AI Systems and help transform public procurement processes. Such an approach will *not* result in concrete applicability of AI Systems without a *coordinated institutional* approach on AI adoption at EU level which can be taken only through a specialized EU Agency (such as EAIA). This will ensure success of this action item via consistency and competence under the same framework.

   The implementation of the objective under Action 6 as currently proposed is fragmented without being coordinated enough at EU level. This may create *operational issues* similar to those which a corporation might face when implementing various tools acquired by different departments without any central coordination with the company's strategy or with respect to the manner in which each tool will fit in the company's ecosystem. Implementing various pre-AI tools with the goal of ultimately adopting real AI technologies should be the roadmap towards the digital transformation of every Member State authority and EAIA would be the right institution for this.

   For this purpose, a unique strategy document should be adopted outlining the principles and framework for every public acquisition in order to ensure consistency and the required technical synergies. Having such a strategy document would ensure that the EU avoid a fragmented and uncoordinated approach regarding the adoption of AI Systems by the public sector ensuring compliance with the AI Principles.

Recommendation 4 – AI Minimum Standard

European Values

The current wording in the AI White Paper only includes a possibility to promote EU values when working and collaborating on AI international projects[6]. However, unequivocal

---

the Internet, and the proliferation of autonomous systems (such as self-driving cars and robots)." Source: http://logic.stanford.edu/publications/genesereth/complaw.pdf

[6] "The Commission is convinced that international cooperation on AI matters must be based on an approach
that promotes the respect of fundamental rights, including human dignity, pluralism, inclusion, nondiscrimination and protection of privacy and personal data and it will strive to export its values across the world."
WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust, p. 9
Source:https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

adherence to a minimum of the EU's legal and technical standards should be a precondition to any possible international cooperation on AI matters. The EU's values should always be expressly reflected in any documents, policies or positions of the EU's institutions. The EU's core values and the rights and liberties of the European citizens need to be clearly stated in any future policy document as a legally binding minimum standard from which the Union will not deviate in its initiatives either internally, in relation to its own Member States, or externally, in relation to third parties, when collaborating on AI projects.

Recommendation 5 – Updating the 7 Key Requirements[7] and the categories impacted by AI as defined by the AI HLEG

New Key Requirements

In line with Recommendation 2 above, the seven principles identified by the AI HLEG in its Guidelines on Trustworthy AI published in April 2019 (the "**Guidelines**"), although non-exhaustive, should strive to be as complete as possible for predictability and proper subsequent regulations purposes. It is important to understand the purpose behind the concept of a "trustworthy AI" and how the development of AI Systems fulfils this purpose.

The AI HLEG stated in its Guidelines that, in order to be trustworthy, any AI System should comply with seven key fundamental rights and ethical principles[8] (the "**7 Key Requirements**"). The HLEG further identifies the different stakeholders involved in the AI Systems life cycle to which the seven key requirements are applicable, categorized as (the "**AI Stakeholders**"):

- developers (those who research, design and/or develop AI Systems);
- deployers (public or private organizations that use AI Systems within their business processes and to offer products and services to others);
- end-users (those engaging with the AI System, directly or indirectly); and
- the broader society (all others that are directly or indirectly affected by AI Systems)[9].

However, stating that all 7 Key Requirements apply to all of the above AI Stakeholders does not take into account their respective role in the life cycle of the AI System. It should be

[7] 1. Human agency and oversight - Including fundamental rights, human agency and human oversight
2. Technical robustness and safety - Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility
3. Privacy and data governance - Including respect for privacy, quality and integrity of data, and access to data
4. Transparency - Including traceability, explainability and communication
5. Diversity, non-discrimination and fairness - Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation
6. Societal and environmental wellbeing - Including sustainability and environmental friendliness, social impact, society and democracy
7. Accountability - Including auditability, minimisation and reporting of negative impact, trade-offs and redress"
Source: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Human%20agency

[8] See note 7 above.

[9] Source: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Human%20agency

made clear which of the 7 Key Requirements each AI Stakeholder should focus on and apply separately. In particular, developers need to know which of the 7 Key Requirements to apply to their design, deployers must verify that the design of the AI Systems they use correspond to the Key Requirements applicable to the developers and that they themselves comply in their use of the system with any Key Requirements applicable to them, while the end users and society in general are always informed about the compliance of the systems they use with all of the 7 Key Requirements (see Fig.1 below for visual representation of the relation).

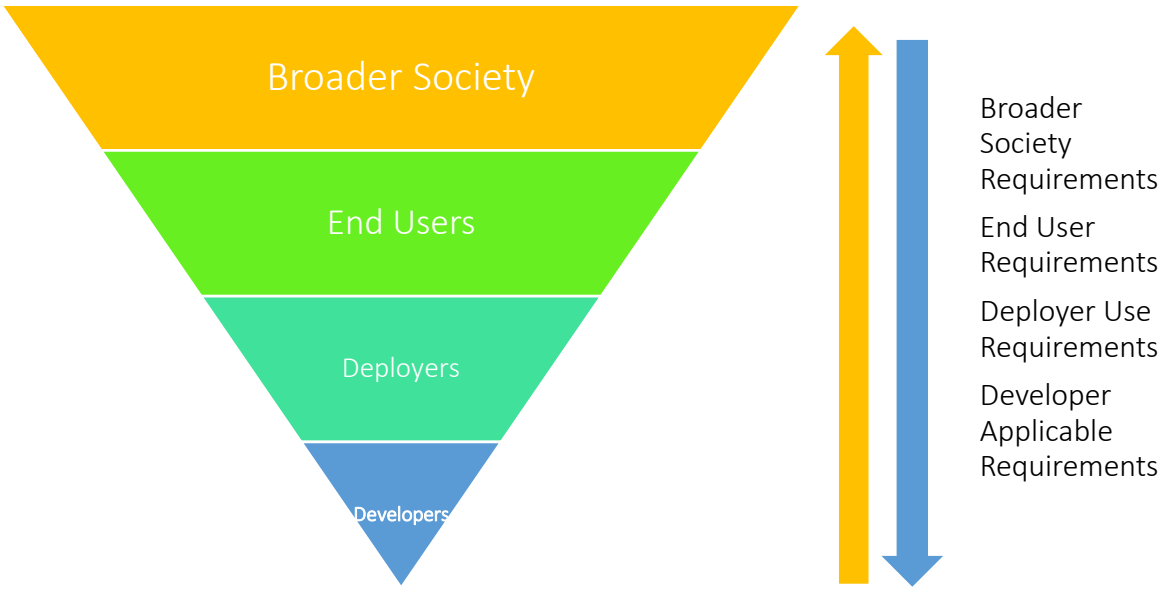

Fig 1. The requirements applicable to each category and the relation of auto-control of the requirements by each specific category highlighting that there should be different responsibility, requirements and liability for the AI Stakeholders. A similar analogy is the separation of powers in a democratic system through legislative, executive and judicial branches, where each branch ensures that the other is not abusing its powers while at the same time each branch has its responsibilities clearly defined. By applying this mechanism, we can ensure clarity over which requirements apply to each category and the extent of the potential liability inherent to a certain category and the stakeholders forming it (joint or separate liability of stakeholders within the same or different categories depending on the requirements applicable to them and their actual role).

It is safe to assume that there are only a number of principles which can be reasonably identified and enforced at this point in the evolution of AI while others will need to be developed once the AI Systems will reach its next stage in evolution and should therefore be corelated with the three AI stages of development by considering the future new principles and anchoring them in these Key Requirements.

To address this view, additional requirements should be added to or included in the 7 Key Requirements, as follows:

1. *Hard coded values.* Similar to the principle of privacy by design, we should identify other human rights or values which might be impacted by unpredictable technological advancements and ensure that these are protected as well.

2. *Prior version reversal.* AI Systems should, by default, allow for the possibility of reversal to a prior uncorrupted version of the system in case of any future voluntary or involuntary breach of hard coded values.
3. *Overriding controls.* The possibility to override the AI System should be mandatory in case of Autonomous or Advanced AI Systems. Whenever a violation of hard coded values or a corruption of the received human mandate occurs, should automatically try to correct it, for example by using the prior version reversal function, or if automatic correction does not work, the system should allow for human intervention. In any case, the AI System should preserve the decision tree.
4. *Learning Parameters*. We would need to establish the right rules for learning and the safeguards and limitations needed to avoid infringing the rights of EU citizens; such rules should always be in compliance with all of the above key principles and requirements; in other words, instead of treating any breach of AI we would be preventing it.
5. *Most Favourable Technique ("**MFT**").* We need to clarify the development techniques and the priority of their use versus the possible impact on the individual's rights.[10]

[10] There are situations where proper guidance could ensure that developers use the right mechanisms to develop AI Systems which technically facilitate legal compliance, such as the work on machine learning models. Certain techniques used by developers would ensure for the regulator a more transparent and explainable view on how the model works which could affect the way liability will apply, in case of any issues. "*While there is no uniformly accepted definition of explainability in machine learning, models can roughly be grouped into those that are interpretable ex ante and those that can be explained only ex post (Lipton 2018; Rudin 2019)*" Reference source: Hacker, P., Krestel, R., Grundmann, S. *et al.* Explainable AI under contract and tort law: legal incentives and technical challenges. *Artif Intell Law* (2020). https://doi.org/10.1007/s10506-020-09260-6
This means that assuming there is a breach of law it would be much more difficult to do a post-mortem to understand how we reached that point, based on the learning model used. In other words, there may be learning models which can be interpreted ex ante while others only ex post. This is important for how the law can regulate the use of these models and include the MFT as a requirement. This would mean that if there is a development scenario in which a developer can choose to apply either an ex ante or an ex post solution to the learning model, the developer should apply the ex-ante solution which can ensure transparency and explainability. "*Linear regression models, for example, are typically interpretable ex ante as the regression coefficients provide an understanding of the respective weights of the features used to make a prediction (Lapuschkin et al. 2019)*". "*Deep neural networks, on the other hand, are typically so complex that specific weights cannot be determined for individual features in a global manner, i.e., for all possible predictions (Lapuschkin et al. 2019; Lipton 2018; Rudin 2019)*" Reference source: Hacker, P., Krestel, R., Grundmann, S. et al. Explainable AI under contract and tort law: legal incentives and technical challenges. Artif Intell Law (2020) https://doi.org/10.1007/s10506-020-09260-6
"*When considering problems that have structured data with meaningful features, there is often no significant difference in performance between more complex classifiers (deep neural networks, boosted decision trees, random forests) and much simpler classifiers (logistic regression, decision lists) after preprocessing.*" Reference source: Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* **1,** 206–215 (2019). https://doi.org/10.1038/s42256-019-0048-x Therefore, a legal preference for more simple and explainable models and techniques to be used when designing machine learning models in similar situations is relevant and required to be provided as a mandatory requirement part of the legal framework – the Most Favourable Technique. The preference in this case related to machine learning interpretable models would ensure that the Transparency Key Requirement (See: 4. *Transparency* -

This means that the development technique which is not negatively impacting individual's rights, transparency of the process, explainability or other Key Requirement applicable to developers, should be used. When there are two or more relatively similar options for development but some of them do not conform with the applicable Key Requirements, they should not be used, if there is the alternative to use one which complies with the applicable Key Requirements. If there is no alternative, then an analysis which would consider all applicable Key Requirements regardless of the AI Stakeholder category will be used to determine which MFT from another category can be used to limit or reduce the possible negative impact.

Nevertheless, all the 7 Key Requirements and the AI Stakeholders to which they apply, as described by the AI HLEG, have as a premise and are focused purely around human actions and their instructions or involvement in the development and design of AI Systems. This human centric premise misses an important purpose of AI regulation, respectively that AI Systems also perform independent actions in all of its stages of development (from Basic to Advanced AI Systems). This is understandable at this point as all the documentation is focused mostly around Basic AI Systems and not on all three types of AI Systems, including those which may be developed in the future. Therefore, the defined principles and categories of people impacted by AI Systems without considering the evolution of AI Systems to Autonomous and Advanced, is essentially flawed because it focuses only on one part of the problem – the human actions.

Recommendation 6 – Additional Stakeholders Impacted by AI

Even at this point, all the categories of human AI Stakeholders described by the AI HLEG can be identified within the AI System itself:

- *human developer* (those who research, design and/or develop AI Systems) – we already have *AI developers* (software which writes software code[11] independently of humans, doing research, design and/or developing AI Systems);

- *deployers* (public or private organizations that use AI Systems within their business processes and to offer products and services to others) – we already have *self-improving software deployers* capable of replicating and improving its own code base by learning from open-source programmers and using the acquired knowledge to make suggestions on how the written code (irrespective of the language or form used) can be improved;[12]

This requirement is closely linked with the principle of explicability and encompasses transparency of elements relevant to an AI System: the data, the system and the business models. Source: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Transparency) is achieved in the development process of AI Systems. Essentially, we need a more complete Transparency key requirement as defined by the AI HLEG which would not only promote explainability but also the choice that would ensure easier and safer the transparency's requirement applicability.

[11] Source: https://news.mit.edu/2019/toward-artificial-intelligence-that-learns-to-write-code-0614

[12] Source: https://analyticsindiamag.com/10-ai-applications-that-can-generate-code-themselves/

- *end-users* (those engaging with the AI System, directly or indirectly) – we already have *software robots' end-users*[13] adopted at a general level by major public and private organizations which execute tasks and workflows in the same manner as a human would; and

- *the broader society* (all others that are directly or indirectly affected by AI Systems) – we already have as highlighted above an *ecosystem of Basic AI Systems*[14] and tools (comprised of *AI Developers*, *self-improving software*, *end-users software robots*) which work towards improving themselves and evolving from what they currently are, which is Basic AI, to Autonomous AI and even Advanced AI.

Therefore, it is imperative not to ignore this reality. Focusing on the actions of the AI System itself will allow us to have a comprehensive approach which takes into consideration both human actions and the AI Systems' actions and the causality relationship among these various actions.

For example, an AI System may initially act following a person's instructions, but subsequently evolve and perform other actions which would breach or infringe the users' rights without it being possible to identify any fault in the human developer or deployer because they wouldn't have been able to reasonably foresee this evolution of the AI System. The cornerstone of the Human Agency Key Requirement[15] – the principle of that user autonomy must be central to the system's functionality – needs to be clarified in light of the above by considering both human designed systems and AI designed systems (either Basic, Autonomous or Advanced). The user autonomy principle, therefore, may be complied with in different forms by the human developed systems versus AI developed systems. We need to understand that depending on the AI's evolution certain things will change gradually in their entirety, including the way in which we think about laws, development, design, language, psychology, ethics, etc. Therefore, the entire approach and strategy with respect to AI needs to take into account things which we may not know or fully understand at this point, but which we could reasonably predict. In the specific case, the *principle that user autonomy is central to the system's functionality* should be redrafted to highlight that *user inviolability needs to be vital for all AI Systems functionalities and versions*. The term inviolability is far more comprehensive than autonomy which is limited only to certain human actions or types

[13] Source: https://www.wsj.com/articles/software-robots-get-smarter-thanks-to-ai-11575887400
[14] Source: https://dzone.com/articles/3-big-trends-shaping-the-ai-ecosystem-right-now
[15] Human agency. Users should be able to make informed autonomous decisions regarding AI Systems. They should be given the knowledge and tools to comprehend and interact with AI Systems to a satisfactory degree and, where possible, be enabled to reasonably self-assess or challenge the system. AI Systems should support individuals in making better, more informed choices in accordance with their goals. AI Systems can sometimes be deployed to shape and influence human behaviour through mechanisms that may be difficult to detect, since they may harness sub-conscious processes, including various forms of unfair manipulation, deception, herding and conditioning, all of which may threaten individual autonomy. The overall principle of user autonomy must be central to the system's functionality. Key to this is the right not to be subject to a decision based solely on automated processing when this produces legal effects on users or similarly significantly affects them.
Source: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Human%20agency

of AI Systems that can influence its decision making process, whereas inviolability would include all the fundamental rights and liberties, including autonomy, that are recognized by the law and which will need to be considered whenever AI Systems will impact any user or the broader society.

Recommendation 7 – Designing the EU's Regulatory Framework for AI Systems

The first elements to be considered for designing any new regulatory framework, including for AI Systems, should typically be: **(i)** the subject matter, **(ii)** the audience, and **(iii)** the risks of having no regulation or an incomplete regulation. Each of these elements allow us to determine the degree of regulation required and the acceptable interference of the competent authorities for ensuring compliance in accordance with EU values and the fundamental rights and liberties of the individual. In the AI White Paper, the Commission is considering creating a regulatory framework for AI around the concept of "high risk" AI Systems.

The Commission is of the opinion that a given AI application should generally be considered high-risk in light of what is at stake, considering whether both the sector and the intended use involve significant risks, in particular from the viewpoint of the protection of safety, consumer rights and fundamental rights.[16] The Commission further indicated that an AI application which meets the two following cumulative criteria should be considered high risk *ab initio*:

(1) it is used in a high-risk sector and
(2) there is a high-risk impact on the affected parties.

Although both criteria are debatable, they would ensure in the Commission's opinion that the scope of the regulatory framework is targeted and clear thus providing legal certainty. Even if at this point the criteria are not finalized in terms of what is the actual expected result, as the stated purpose is too unclear to be relevant for effective AI Systems regulation framework, it is essential to clarify them. Their clarity and effectiveness will legally set the EU's AI Systems for success or failure because the AI Systems identified as being high risk based on this assessment alone will trigger certain regulatory requirements and oversight for those applications by default.

It is also stated that the criteria and their applicability is fluid and other non-high risk determined applications in "safe" sectors might become nevertheless subject to the high-risk status as long as they might have an impact of individuals. Due to this carveout the entire high-risk assessment versus "safe" applications loses its value as it is achieving exactly the opposite of what the regulator wanted to achieve, respectively clear framework and legal certainty. To ensure that this objective is achieved, the Commission should remove this mechanism stated in the AI White Paper which will not only create uncertainty but might also

[16] WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust, p. 17 Source:https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

increase bureaucratic burden on the EU institutions and agencies and national authorities when trying to assess if a particular AI application should be catalogued as high-risk (human and AI activities will constantly evolve and change which will make it highly difficult to keep track of everything created and worse it can act as a show stopper for important technological advancements due to the need for regulatory validation).

Therefore, instead of using this risk criteria assessment, which is discretionary by placing certain sectors as high-risk in their entirety but also has the generic exception which can make anything subject to it, thus regulating uncertainty, the Commission should focus on simply clarifying what it wants to safeguard and protect. It should, therefore, indicate a very *clear set of instructions* related to the values it wants to protect leaving it to the developers and deployers to find creative means to ensure compliance with the EU's values and its citizens' fundamental rights and liberties while also continuing to innovate in their respective fields.

A possible solution could be defining an *evolving system of rights and values* which can be clearly applied to the development process similar to the Global Data Protection Regulation's *privacy by design principle* when developing new solutions or *the right to be forgotten* when an existing solution needs to be capable to remove certain information from public access. The human and AI developers need to easily understand *what* and *why* the regulator is requiring to be considered at the development stage but also post-development. In this way, the system can be focused in "hard coding" the 7 Key Requirements to become legally and technically applicable principles in the development of any AI System, irrespective of the sector or how it will be used. As long as an AI System has in some form or another the Key Requirements embedded during the development or post-development process, it should be able to exist without the need for any further special or extra regulations.

The Key Requirements need to be translated in technical capabilities, as follows:

- *Human oversight* = privacy by design, possibility of choice, accurate information and source, data quality, possibility of rejection, deletion, change, etc.;
- *Technical safety* = overriding & security controls, accuracy & reliability;
- *Transparency & Accountability* = explainability & traceability, audit;
- *Non-discrimination* = accurate data & unfair manipulation of data;
- *Environmental impact* = sustainability via low resources use & saving energy;
- *Prior Version Reversal* = reversal by default to a prior uncorrupted version of an AI System;
- *Most Favourable Technique* = establishing the right techniques used when creating AI Systems.
- *Learning Parameters* = establishing the right rules used by an AI System to improve.

Recommendation 8 – AI Complexity and Novelty: The need for an Artificial Intelligence Regulation (A.I.R.)

Separating between AI Systems Development and Liability

The AI White Paper is approaching the subject of liability for an AI System or of the AI System, without first having a proper *understanding* of an AI System and its different stages of life development (pre-AI Systems, Basic AI, Autonomous AI, Advanced AI). Such understanding is a pre-requisite before working towards regulating the AI Systems, as part of the *preparation* phase, and only after such preparation will we be able to design the right *safeguards,* which if not complied will put liability into question.

The issue of liability is approached somehow fragmentary with a generic indication of already existing laws and initiatives which indicate that current approach is oriented towards a minimization of regulation and re-purposing of whatever each Member State or authority already has in place related to the subject or is thinking to do in terms of AI regulations and its impact on the current legal framework. Furthermore, the AI White Paper is not entirely clear in its approach on this matter. For instance, it states that there is sufficient regulation to sanction any breach of existing rules by AI Systems and that economic actors remain fully responsible.[17] This is inaccurate as liability of economic actors can be widely different based on their role in a specific liability scenario which would be left for the courts to decide without having a complete legal framework. For example, would the developer of the software be liable for the way in which a piece of software is trained and subsequently is used by deployer or a user?

A more productive approach would be instead to proactively identify the right premises to understand *what* needs to be regulated, *how* it will make more sense, for *which* consequences we should prepare and *who* will apply such regulations.

If an AI System is not inherently designed, nor supposed to be used for military purposes, but the users can nevertheless use it for such purposes (due to the fact that it is an AI System which is able to be used or is expected to be used as a piece of software capable of improving itself), should it be catalogued by default as a dual use software and thus apply to it *ab initio* a more strict legal framework? The same question can be asked in case of environmental, medical or nuclear devices. In this case, the best way to approach the liability issue or more broadly the way they are regulated, would be under the assumption that any AI Systems may evolve or have different uses then the ones first intended. Therefore, not only the liability of the developer should be brought into question, but also of the deployers and even end users, depending on each of their roles in the final use of the AI System.

Nevertheless, there may be AI Systems for which developers will be able to confirm that they will remain in the same stage (Basic AI or Autonomous AI) and that their use cannot be changed. For example, in case of self-driving cars, we could identify all three different types of software: (i) a Basic AI software which is limited to analysing the surrounding environment and transmitting the data to the second type of software, (ii) the Autonomous AI that can

[17] For example, economic actors remain fully responsible for the compliance of AI to existing rules that protects consumers, any algorithmic exploitation of consumer behaviour in violation of existing rules shall be not permitted and violations shall be accordingly punished. WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust, p. 14
Source:https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

extract, interpret and use that data to drive the car, while an (iii) Advanced AI would be capable to monitor and interact with the main code of every self-driving car and decide which to stop or re-route based on certain parameters. In this case, given that the Basic AI System used in such cars cannot transform into the following types on its own, the liability for such system would, in principle, remain with its developer.

The AI White Paper also identified to some extent that there is a need to clarify and improve the scope of certain EU legal instruments to proper address AI evolution as the current existing horizontal and sectoral legislation is not sufficient. Nevertheless, the Commission's approach seems to be in testing the waters mode and there are no decisive measures highlighted or expected to be taken following the AI White Paper.

Therefore, in line with the above recommendations, the EU should look at adopting a regulation on artificial intelligence which would have direct applicability across the EU. This will reduce the possibility of misalignment or inconsistencies among the EU's and its Member States' approach in this matter. Such regulation would cover all the requirements needed for properly preparing for the AI evolution and its impact in society. Although AI is currently in its infancy, it has an unprecedented opportunity to reach unpredictable evolution which will affect each and every one of us as well as our institutions. Regulators need to understand this opportunity and focus on what matters by approaching the subject of AI regulation with the right urgency and with the AI specific requirements in mind which will ultimately guide the way society will continue to live.

Summarizing, there is a strong need for a clear regulation establishing an EU focused Agency on AI, complete the proposed 7 Key Requirements for trustworthy AI, clear responsibility, requirements and expectations from AI Stakeholders and the extent of their liability in this process. Only by consolidating these efforts under a common institutional framework will the EU overcome its current status and will assume the much-needed ownership.

Conclusions

The European Union has a choice to be either bold and transformative or complacent and expectative. In the first scenario EU would take an aggressive approach towards reducing the existing gap with its main counterparts, such as US and China, and take the lead in capitalizing its existing potential. Or it can take the second option and maintain its international status quo of a slow responding and bureaucratic entity which will continue to be a top provider of talent and technology to US and other international markets and fall in line with the direction these markets will set in terms of AI development, evolution and adoption.

The stakes are very high and could lead to a shift in the global markets if the EU would go with the first scenario. According to McKinsey Global Institute, if Europe on average develops and diffuses AI according to its current assets and digital position relative to the world, it could add some €2.7 trillion, or 20 percent, to its combined economic output by 2030. If Europe were to catch up with the US AI frontier, a total of €3.6 trillion could be added to

collective GDP in this period.[18] This finding is based on an average estimated effort, while if the EU would take an aggressive stance for designing and executing an AI strategy that would be truly transformative for its single digital market. It would also represent a unique achievement considering the EU's challenges in providing a single viable framework for its Member States in various fields of cooperation and EU integration compared to US and China which have less internal complex challenges from an institutional point of view.

The current technological and economical gaps can be overcome only with innovation which needs to start from an agile legal framework up to lean institutional mechanisms which will set the right premises for technical development and a trustworthy European AI ecosystem.

[18] https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-europes-gap-in-digital-and-ai