# Response to EU Whitepaper on AI

## appliedAI Initiative of UnternehmerTUM

## General remarks to the Whitepaper

The whitepaper correctly emphasizes the need for Europe to become a global leader in AI. While the Commission pushes for a value-based and principle-driven approach, we still need to earn our 'right to play' in the world class arena. Europe must have a significant impact in AI development and application in order to have a real say in the global development. Otherwise we lose competitive power while keeping theoretical principles high. The whitepaper points out two ecosystems to achieve the goal of becoming a global leader. We welcome the outlined ambition and the consolidated approach emphasizing the importance of a single digital market, and the willingness to invest significantly in AI.

The ecosystem of excellence highlights measures to lead in research, skills and innovation. However, for innovation and application we suggest several additional measures that contribute towards achieving the goal.

The ecosystems of trust lays out measures to increase trust. However, if regulation is to be emphasized, the opportunity cost of not using AI and the additional benefits that might come through the application of AI technologies should be mentioned and taken into consideration significantly. For example, if we talk about autonomous driving, we should recognize the drastically increased safety compared to human drivers and not only the potential risk of an AI induced accident. Similar benefits could already be observed in the use of AI in China in fighting Corona via Telemedicine and Robocalls on a massive scale, as well as autonomous cleaning and delivery in infected areas (reference). Moreover, it promotes a technology specific risk-based approach to a technology that is difficult to define.

There is an imbalance within both parts in regard to regulation being the number one answer for building trust. This is not supported by evidence when international attitudes towards technology are analyzed.

Both the ecosystem of excellence and the ecosystem of trust scarcely consider speed and agility, although the first sentence of the whitepaper "Artificial Intelligence is developing fast" acknowledges the importance of this aspect. We would welcome a greater emphasis on this aspect in all proposed measures or even going further and adding this as a third pillar - arguably a top priority at this stage. This is particularly critical in view of the peculiarities of European decision making: It is comparatively easy to agree on an initial joint framework, since the value of commonality is so high. These dynamics is absent when changing existing legislation which will be required regularly for ever changing AI based systems and services. Finding methods to dynamically adjust to the advancements of AI is mandatory if we follow a regulatory approach. The EC should be attentive to regulatory developments in other regions of the world to make sure that EU companies are not slowed down and therefore disadvantaged by new EU regulations.

Overall, the paper takes a reactive and conservative perspective rather than encouraging Europe to take an influencing and driving role. However, this is essential for European companies to be able to compete on the global stage and thus for the EU to truly shape the developments. The Commission needs to draw a positive picture on AI application to promote interest and acceptance of this powerful technology and to avoid a too risk-focused perspective. Measures should be more target-oriented and visionary. Regulations should only be put in place when needed and agility and speed need to be the north star. Only then can Europe play out its strengths to keep up globally.

## 1. Ecosystem of Excellence: The ambition vs. outlined measures

*"The Commission is committed to enabling scientific breakthrough, to preserving the EU's technological leadership and to ensuring that new technologies are at the service of all Europeans – improving their lives while respecting their rights"* (Introduction, p.1).

Improving lives is only achieved through innovation and application and not on a theoretical level. The actions outlined in the whitepaper will not be sufficient to achieve the goal of global leadership. The whitepaper takes a single European perspective ("*The European approach for AI aims to promote [...] across the EU economy*", p.25) but it must be embedded in global activities, target the application of European trustworthy AI also outside of Europe, and foresee the actions and reactions of other AI countries. Missing from the paper is a strategy which builds on Europe's strength while reducing its weaknesses. Especially, when it comes to startup and innovation activities more activities should be taken. Moreover, many actions suggested in the paper may even slow Europe down in relation to other parts of the world. Currently, when compared to China or the US - or even Israel and Canada - Europe does not create new global leaders in AI. None of these regions strive for parochial 'regional leadership', they all compete globally. It may be worth considering to outline and monitor KIPs for "global leadership" to better focus on the most effective actions.

In addition to the actions mentioned in this chapter, the ecosystem of trust should also reflect the mindset of the ecosystem of excellence. Regulation should encourage innovation and support the goal of global leadership. Excellence in regulation would mean close coordination within a common European framework. A "*forum for a regular exchange of information and best practice identifying emerging trends, advising on standardisation activity as well as on certification*" with a "*cooperation of national competent authorities*" as outlined on p.24 describes a cumbersome and slow process. It is inadequate to reach global leadership and avoid the fragmentation of the internal market. We would highly welcome a consistent excellence-driven European approach (a "European AI Core") built on standards, norms, and certification.

## 1.1. Action 1: Working with member states: (p.5)

We welcome the plan to invest 20 bn EUR per year in order to remain competitive globally even though it remains unclear if this investment comes on top of existing funds or if it consists mainly of already ongoing activities. In light of Covid19 and the challenges that come with climate change, it should be outlined that AI helps address these challenges - we will not manage without it - and thus has considerable benefits for the whole of society beyond the individual perspective.

## 1.2. Action 2: Focusing the efforts of the research and innovation communication: (p.6)

The Commission correctly points out that "*Europe cannot afford to maintain the current fragmented landscape of centres of competence with none reaching the scale necessary to compete with the leading institutes globally*". Further effort and most importantly speed and long term, outcome-oriented commitment must be put into creating centers of excellence in order to avoid member states creating a chain of repetitive, subscale activities. There should be a clear statement that the Commission will move ahead with willing member states. This will help to avoid becoming blocked while waiting for all laggards to join.

It does not become clear if a lighthouse centre of research is considered a single centre or a virtual centre consisting of many existing organizations that have joined forces. The current activities (e.g. ICT-48, ICT-26 calls) point to a networked approach. We would welcome a well-orchestrated network. The creation of networked lighthouses and centres requires communication and rigorous quality assessment to support dissemination and knowledge transfer. An approach like the DARPA challenges including substantial funding, that is proven to be outcome-oriented while reaching high quality, might be considered to maintain competitive forces throughout the operation of these centres. Moonshot activities (e.g. lighthouse cluster on trust, the self-driving EU car, European AI center against climate change) along with close ties to the Commission might also be used to attract top experts and talents.

With regards to the test centres as well as the lighthouses, it is not clear if they are research or application driven. We would strongly encourage a very application driven approach.

### 1.3. Action 3: Advanced Skills (p. 6)

We welcome the described measures to increase capabilities and to attract talent to Europe. The skills activities should be aligned with the overall goal of the Commission. It should encompass not only application fields but also address less attractive research directions like liability, testing methods, or transparency.

### 1.4. Action 4: Focus on SMEs (p. 7)

While we support the actions taken to support SMEs and the startup ecosystem, the outlined measures do not suffice. Strategic European Champions in the startup sector need to compete globally with teams that receive significant public long term contracts (e.g. SpaceX/NASA, Sensetime/Chinese Cities). Besides a financing pillar which needs to be well beyond 10bn EUR, tender procedures need to be adjusted to allow for the creation of new global champions. This might be tied with the lighthouse centers and a setting like the DARPA challenges. Digital Innovation Hubs need to be substantially funded to share high quality knowledge and assets and to network to be able to support SMEs sufficiently. This does not go without saying that innovation is not dependent on the size of the company (start-up, SME, large corporate). Preserving a level playing field should not be disregarded.

### 1.5. Action 5: Partnership with the private sector: (p.7)

The significance of a European data strategy as an essential foundation for AI is recognized and fully supported, along with the important role of international cooperation, standardization, harmonization and mutual recognition of standards and a regulatory structure.

The private sector acts in a globalized world. While our corporate partners welcome partnering on AI with the Commission, it needs to be emphasized that ultimately AI will become a competitive factor in the world economy and a global perspective is paramount for global players. Everyone should be careful to look at the topic from an isolated European standpoint not taking into consideration how the rest of the world is approaching AI. A specific focus should be set on the topic of how we can maintain high speed (as can be seen in other parts of the world) in a values-based approach to AI.

To directly contribute to the European AI agenda, a feedback and suggestion mechanism from the private sector could be established to create an application focused input for the European Research Agenda.

### 1.6. Action 6: Promoting the adoption of AI by the public sector: (p.8)

The “Adopt AI” programme as outlined in the whitepaper is a very important component for the uptake of AI in Europe. We would welcome an ambitious budget.

## 2. Ecosystem of Trust: The ambition vs. outlined measures

“*Given how fast AI is evolving, the regulatory framework must leave room to cater for further developments*" (p.10). The ecosystem of trust rightfully points out the speed of development. Yet, it falls short on connecting the proposed measures with this most central sentence. The section gives the impression that trust in the technology can be achieved through regulation and certification only (*[...] a clear European regulatory framework would build trust among consumers and businesses in AI, and therefore speed up the uptake of the technology.* (p.9)). Yet, trust can also be built through technology, standards, or market-driven approaches. Regulation should only be applied when needed and should avoid addressing aspects that are software-specific and not limited to AI. A principles-based global framework on Data Ethics and AI could be beneficial to reflect common understanding of the relevant existing legislation. Companies might either sign a public statement to comply with the framework or they get certified. In order to avoid market fragmentation and impediments to innovation any EU framework should reflect global principles adding more particular requirements only if necessary to prevent harm. It needs to be recognized that AI is currently in development with no established framework for certification, ongoing progress on explainability, the industry-driven development of standards to mitigate bias and massive research happening globally. The ecosystem of trust and the outlined regulation must reflect this dynamic environment to avoid that major developments and application will happen outside Europe. Regulation must cater to rapid changes, anticipating future developments not having the status-quo as basis. Also, the risk-based approach needs to consider ongoing reassessment. National bodies need to be able to follow the advances in AI constantly. Failing to do this would mean to fall behind more innovation-friendly environments in China, the US, Singapore and other countries. Given this challenge, we would welcome a competence center in the EU guiding and supporting standardization efforts, consulting Member States and the industry on regulatory measures and constantly monitoring technological developments and end-to-end effects of existing legislation.

Overall, we welcome the guidelines for trustworthy AI and a risk-based approach as there is no “one-size-fits-all” solution for the multitude of applications that are affected by AI. However, the Ecosystem of Trust describes a very defensive picture of handling AI which might make it very hard if not impossible to achieve the goal of global leadership **in practice.** Therefore, before regulating, the EC should assess the impact on AI innovation and growth and take an active role in guiding the development. Following that, we mandate for significant adjustments of the ecosystem of trust, summarized in the following topics;

- General remarks on trustworthy AI (p.9)
- Liability (p.12-15)
- Standardization before regulation
- Risk-based approach (p.17)
- Specific requirements for high-risk cases (p.18-22)
- Effects on sharing / open source

- Monitoring (p.23-24)

### 2.1. General remarks on trustworthy AI (p.9)

The guidelines for trustworthy AI of the AI HLEG define principles that should be followed in the EU. However, the application of the principles to any AI use case might be very case-specific and not limited to AI. Therefore, we see it necessary to comment on several principles.

Bias, discrimination and fairness: On the one hand bias and discrimination may or may not be challenges of AI technologies. Bias is a well established concept in data science education and thus a methods for handling bias are available. If bias is unwanted and must be reduced in a specific use case, bias reduction is by design in the process of building AI applications. On the other hand, humans routinely make biased decisions. AI is measured much more rigorously and can be re-engineered instantly, so it may be better in reducing bias than managing humans could ever be. Although the data on which an algorithm is based still plays a significant role, potential discrimination only occurs when the trained algorithm is applied and specific criteria can be imposed on the results. Any potential regulation should take this into account and not create major additional obstacles for the recording and quality of training data in itself. Rather the regulation should word the requirements in such a way that any potential discrimination in the selection of data and its use in training algorithms is sufficiently considered. In fact, discrimination has been covered by law for a long time and we do not see any convincing argument that existing law is not sufficient. While bias and discrimination can be handled reasonably, fairness (outside a definition of "significant influence of a random irrelevant variable on the result") is a very difficult concept and this should be considered while deciding whether to include any related requirements as part of regulating AI applications.

Explainability and Transparency: The AI-specific challenges of explainability (and transparency) are of inherent technical nature. Significant research is conducted to resolve these on a scientific level and solutions are highly relevant for our industry in

their own interest. Findings and technical solutions should be translated into sector neutral standards without significant regulation. Following scientific progress, a reasonable time for implementation on existing AI applications must be given.
Instead of new regulation, we fully support the extension of existing transparency rules to customers with alternative solutions that provide equivalent customer benefit such as a customer's right to ask for human (re-) validation of the correctness of the algorithm-generated result.

### 2.2. Liability (p.12-15)

The existing – technological neutral – liability regime is quite comprehensive and should be made applicable to AI with necessary clarifications before new concepts are introduced.
A separate Civil liability regime for AI as suggested by the EU Parliament might hinder innovation and be counter-productive, as it introduces strict liability for high-risk applications in the public sphere. Additionally, it suggests fault liability for non-high-risk applications beyond contractual relationships next to the existing EU civil liability regime.
The existing EU Product Liability Directive (EU 85/374) should be amended to also provide guidance on the matter of liability for embedded software including AI based (self-learning) algorithms and applications. Final documentation and duty of information under the PLD should be equally applicable to any AI applications that can have an impact on customers and citizens, irrespective of the assumed qualification of high risk. However, any new compulsory requirements like ex-ante testing and approval by authorities should be limited to high-risk applications only. The PLD should include ongoing monitoring and updating obligations to the developer and deployer of AI based products.

### 2.3. Standardization before regulation

The commission rightfully points out the relevance of trust. Trust is also in the interest of the market and each active participant in order to broaden the acceptance and application of AI technologies. Therefore, standardization and certification activities for adhering to methods and procedures are of highest priority for the market. Regulation should only be applied if market forces are not expected to achieve the principles outlined in the whitepaper. Playbooks for the use of AI and interpretations of existing legislation may be faster and more targeted than new regulatory activities. The Commission should prioritise clarification and guidance on existing legislation before creating new legislation.

### 2.4. Risk-based approach (p.17)

The risk-based approach is welcomed in principle. Yet, there are several aspects we want to highlight for reconsideration in order to ensure that any potential regulation is targeted at the right use cases, provides legal certainty, and does not discourage the development and diffusion of AI.

- Risk: The outlined approach seems to define risk without considering the cost of an alternative option or potential good of the AI solution. Even though the use of AI might involve risk, there might be more harm if AI is not used. Cancer detection may be wrong in 5% of the cases but if an average physician has a 20% probability for misdiagnosis, the AI solution might be preferred. Similarly autonomous cars are likely to cause far less fatal accidents than humans while the risk of causing one is still there. Also, as demonstrated in fighting Covid 19, AI's speed and scale can be of tremendous advantage in saving lives. We propose a balanced risk assessment with negative and positive effects both being considered in the classification process.
- Two classes: In the proposal of the German Data Ethics commission, there were five levels to allow a more differentiated view.. We would welcome at least three classes (high, medium, low risk) as this has several advantages. There is some risk involved in any application and some may need regulation. Yet, only few need drastic external involvement. A two classes system has the possibility of many less risky applications being classified as high risk because the lowest risk class seems too relaxed. Moreover, three classes would allow higher flexibility and could prevent the high-risk class from becoming the catch basin for more and more applications with any little risk.
- Sector based classification: The sector classification seems to be unnecessary. In every sector AI could be used for safety-critical applications as well as purely supportive functions. Thus, the Commission should avoid classifying whole sectors as high risk. We would propose a technology and system-/application based classification model which follows existing sectoral regulation. Moreover, the system that consists of more than just the AI technology and its intended application should be the decisive criteria. This would recognize the varying relevance of AI technology in whole systems (backup, recommendation, autonomous decision). Any additional regulation should be within existing sector-specific frameworks which in many fields seem to be sufficient or very strong already (e.g. for autonomous driving, healthcare)
- Measurement: The risk classification should be formulated to prevent legal uncertainty and allow for self-assessment. Therefore, it should precisely define the classification process including white lists of exemplary cases / classification reasons for each class. The target should be that every company (especially SMEs) is able to assess without external support.
- Probability assessment: Regulation of AI applications must take the changing risk of any AI solution along their life-cycle into account. The intended use as well as the probability of some manifesting risk changes throughout the lifetime of an AI

system and seems very hard (without proper technical support) to predict and to control.

### 2.5. Requirements for high risk cases (p.18-22)

The Commission describes specific requirements for high risk cases. In general, we propose that most of those requirements need to be amended to avoid hindering innovation of European companies. The requirements should focus only on the objective, holding the companies accountable for the operationalization. Otherwise, this may lead to outdated requirements, inconsistent or even contradictory rules and uncertainty in the application.

Training data: There is too much emphasis on training data quality reflecting a focus on past standard supervised learning from labeled data, not on future AI technologies. Data augmentation, transfer learning, generative adversarial methods or even model-based reinforcement learning approaches will prove elusive. Also, a high quality of training data is in the core interest of the company. Moreover, any rules for transparency beyond the existing regulation (e.g. GDPR) might affect IPR and trade secrets of a company and should be avoided. More importantly from a regulator's perspective are standards for test data and testing environments to assess the quality of an AI application. It must be noted, however, that due to the nature of AI systems, it is not possible to conduct tests of 100% of all possible scenarios.

Data and record-keeping: We welcome a clarification regarding the documentation and retention obligation for development documentation. A general keeping of datasets should however not be mandatory. On the one hand, this is likely to conflict with GDPR provisions requiring deletion of personal data. On the other hand it conflicts with copyrighted datasets authorised for only short-term access (e.g. one-year license for input data allows to use the trained model afterwards but not keeping the data). Any change to the training data would make reproduction impossible. Moreover, it would destroy the privacy benefits of on-device processing because it would effectively force data to be collected and stored centrally. Ultimately, it also conflicts with the targets of the Green Deal as it would consume significant resources to store the data sets. Therefore, we strongly recommend that any decision about storing/deleting data (except for limited cases) should be left to the companies.

Robustness and accuracy: It should be well accepted (and understood) that AI will make mistakes and 100% accuracy is not possible. No developer can possibly ensure the level of accuracy during all life cycle phases due to the nature of trained models based on historic data in an ever changing world. We propose scenario-based assessments for high risk cases following the practices in the financial industry and validation mechanisms for automated driving.

Human oversight: When considering automation, AI based systems should be used in situations where this presents an improvement to how a human would perform  (e.g. decisions in split-seconds or highly-complex situations). Thus, human oversight is only applicable in limited situations and following the interpretation of human oversight as monitoring and reaction but not upfront clearance of each decision. If AI based systems are designed for augmenting humans in their decision process (e.g. giving radiologists recommendations), human oversight is given.

### 2.6. Effects on sharing / open source

AI lives from a vivid open-source ecosystem in which training data sets, pre-trained models, or network architectures are shared within the community. The proposals in the ecosystem of trust and more specifically the requirements for high-risk cases would limit or even eliminate the open source ecosystem. Strict liability rules or requirements for data storage and documentation that falls back to the developer of open-sourced data sets would make it impossible to share. In addition, no developer could use pre-trained models (trained generic models for e.g. basic language understanding that someone makes available to be further specified through own data; transfer learning) if the original training data is not published as well. This contradicts the principles that the Commission outlined in the white paper as our European values as well as the targets in the ecosystem of excellence. Therefore, every regulation should be assessed against unwanted effects on sharing, open-sourcing and cross-company collaboration or even be evaluated on the improvement of the mentioned aspects. Checks should be applied to the results of AI based systems, not to the input.

### 2.7. Monitoring (p.23-24)

If new conformity assessments for high-risk AI applications in non-harmonised sectors become necessary, we propose a two-step approach: 1) to perform ex-ante self-assessment against agreed international standards, coupled with ex-post market surveillance; 2) the EC or a supervisor would monitor and evaluate the application of this framework to determine the need for modifications in the light of technological or market developments. Any ex post testing should be proportionate to the level of risk of the AI based application.

In general the ongoing testing and alerting along the whole life cycle will become more important than an upfront testing. It could include “stability over time”, “scenario-based testing”, “benchmarking against a standard proprietary test set”, explainability tests (e.g. feature relevance”)

## 3. Additional Comments

Data:

Data is a resource that can be used for various applications. Therefore, it is particularly complex to define requirements for data whose suitability depends significantly upon its use in a specific product or service. Legally binding regulated guidelines on data quality requirements would need to be described, if at all, in the context of their real- world application. As data can generally be used in or to train multiple different applications, it will be difficult to stipulate absolute data requirements beyond minimum standards for data quality, completeness, and representativeness. It is important to state clear purpose limitations and meaningful metadata to describe the data sets, as a basis for making and documenting a choice for product development. However, this is already covered in the extensive regulation of product liability, quality, safety and reliability.

Data Sharing:

We believe that a liberal data economy fostering fair access to and free flow of data whilst protecting investments and trade secrets enhances innovation and thus leads to better services and products for European citizens. Yet, specific value-characteristics of data need to be considered:

- time-value (e.g. real-time data is almost useless if provided late)
- context-specific value (e.g. a search item is more valuable if the location or situation is known. Provided without context (e.g. anonymized), the value is marginal)
- explicit value (e.g. a data set with a raw data and a label providing the information about the data)
- knowledge value (e.g. data is the information like a protein-structure)

Any regulation that tries to enforce data sharing should be aware of these different aspects and the value creation that lies in the data collection. While public (tax-payers) data should be shared (anonymized or not anonymized) for dedicated use cases, company data needs to be handled sensitively due to the mostly unknown value as well as unclear access rights. An obligation to share data, which has been enhanced, enriched or aggregated (i.e. business secrets) with other industries or competitors must be avoided as it would disclose core business strategies. In our view, a set of minimum requirements regarding access to, and portability of data within the EU such as standardization and interoperability would be helpful to enhance the free flow of data. The Commission aims for global leadership, but global leaders need to keep valuable data proprietary.