August 2021

Microsoft's Response to the European Commission's Consultation on the Artificial Intelligence Act

Introduction

Microsoft applauds the Commission for its leadership in developing a regulatory framework for the responsible development and use of artificial intelligence technologies ("AI"). The Commission's proposed Regulation for harmonized rules on AI (the "AI Act"), and the New Coordinated Plan on AI, are ambitious and important steps toward making trustworthy AI the norm in Europe and around the world. We share the Commission's goal to ensure that the vast potential of AI can be realized by all in ways that are safe, respectful of fundamental rights, and align with European values. We are committed to supporting this effort as part of our 'Tech Fit 4 Europe' initiative to help Europe realize its AI and other tech-related aspirations, and we appreciate the opportunities made available by the Commission to participate in the preparatory work for the AI Act.

We support the AI Act's vision and direction, in particular that it adopts a risk-based framework focused on a clear list of high-risk uses as well as certain prohibited uses, recognizes the benefits of transparency in promoting trustworthy AI, provides for self-assessment in appropriate cases, and incentivizes the broad adoption of responsible AI governance through codes of conduct. We urge policymakers to preserve these elements as the AI Act moves through the legislative process.

Microsoft also believes there are ways to strengthen the proposal. We offer some suggestions based on our experience of developing our own internal responsible AI program and the emerging trends that we see in the commercial and research landscape. These suggestions are informed by the many innovative ways in which our European partners and customers are deploying AI. Our colleagues at LinkedIn have also offered their thoughts by way of a supplementary submission.

While we firmly agree that the AI Act should articulate **what** regulated actors should seek to achieve, we believe it will provide a more effective framework for responsible innovation – now and into the future – if it is less prescriptive on **how** they achieve them. A more principles-and results-based approach that is calibrated specifically for the AI ecosystem will help to drive clarity for regulated actors and spur innovation within necessary guardrails. With these objectives in mind, we urge policymakers to strengthen the proposal in three key areas:

> **Calibrate regulatory obligations to better reflect the roles of different actors in the AI ecosystem.** To the extent AI systems present risks to fundamental rights, these tend to be sociotechnical in nature: they concern the capabilities and limitations of the technology, combined with people's expectations of it and the societal context of its use. When this sociotechnical dimension is combined with key features of the AI ecosystem, including the availability of general-purpose AI services that customers put to use in scenarios of their choosing, the customizable and dynamic nature of AI systems, and their increasingly general capabilities, it is evident that addressing

fundamental rights risks for AI systems requires a different allocation of responsibilities than that designed for product safety risks in the New Legislative Framework. We offer several suggestions for how the AI Act can more appropriately allocate obligations to the entity that is best placed to identify and mitigate the relevant risk.

Adjust the AI Act's requirements to focus on outcomes and processes. As currently drafted, many of the AI Act's requirements are prescriptive or focused on specific scenarios. As a result, they are likely to be unworkable in some cases, ineffective in others, and are neither future-proof nor innovation friendly. A more effective approach would be for the AI Act to clearly articulate the **outcomes** that regulated actors should seek to achieve, together with a list of key **processes** they must adopt to achieve them. This would help ensure that regulated actors identify and mitigate risks in an effective and efficient way across the full range of different AI systems and use cases covered by the AI Act. It also creates space for innovation borne out of future technological developments and advances in the practice of responsible AI.

Adapt the post-market monitoring regime to better address the sociotechnical nature of many AI risks. As drafted, the AI Act's post-market monitoring obligations appear to fall almost entirely on providers. Although apparently modeled on the EU's product safety framework, this approach is not well suited to AI services, since suppliers of AI services often will have no visibility into how their customers deploy these services into their own systems, and in many cases would need to work with customers to address any risks that might arise in specific deployments. We offer several suggestions for modifying the AI Act's post-marketing monitoring rules to better address the realities of the AI ecosystem today and into the future.

We conclude by setting out further changes to promote innovation and strengthen fundamental rights protections for affected individuals, including in relation to law enforcement use of remote biometric identification technology.

Substantive comments

1. Calibrate regulatory obligations to better reflect the roles of different actors in the AI ecosystem

AI systems can create risks to fundamental rights that are **sociotechnical** in nature: they concern the capabilities and limitations of the technology, combined with people's expectations of it and the societal context of its use. For example, risks may emerge at the intersection of system design decisions taken by AI model builders and decisions taken by deploying organizations as to how and where and when to use the AI model in a final system.

Mitigating the sociotechnical nature of AI risks requires cooperation between the different entities in the AI value chain. In particular, the supplier of a general-purpose AI service, with its knowledge of the decisions taken in designing and developing the technology, should cooperate with the deployer so that the deployer can make informed decisions about whether the technology is fit-for-purpose in their chosen use case and how to identify and mitigate attendant risks.

The grounding of the AI Act in the New Legislative Framework (NLF) creates challenges in allocating obligations across the AI value chain. This is particularly true with regard to risks to fundamental rights, which will most commonly arise with respect to the high-risk systems listed in Annex III. The NLF uses pre-market controls to address the health and safety risks that products can pose. The responsibility for meeting the NLF's essential requirements is placed on the 'manufacturer' as the entity that assembles the regulated product from its component parts and controls the design and production processes. 'Users' are subject to very limited obligations, typically in the nature of following assembly or installation instructions. This allocation of responsibilities makes sense for product safety risks for physical products: those risks are relatively consistent across deployment scenarios and seldom change once a product is placed on the market. By contrast, a similar allocation of responsibilities is not well-suited to AI systems, given the general-purpose nature of many AI models and the significant impact that context of use has on the fundamental rights risks that the AI Act seeks to mitigate.

Indeed, AI systems are fundamentally different from the physical products currently subject to the NLF in a number of important ways:

AI models are often made available as general-purpose AI services that are put to use in customer-defined applications

Microsoft offers AI technologies to customers in a number of ways, including embedded in first party products, such as PowerPoint, where our speech-to-text transcription technology provides simultaneous transcription of a presenter's remarks. However, Microsoft and other leading providers also make available their AI technologies in the form of general-purpose AI services. Microsoft's Azure Cognitive Services, for example, comprise a set of pre-trained AI models made available as a service that customers access on a subscription basis via an API. Each of our Cognitive Services performs a generally applicable function, such as text analytics or anomaly detection. Customers decide how and where to embed this functionality into their operations, significantly impacting the intended use of the AI technology and its related risks. For example, a restaurant might choose to use our text analytics service to scan large numbers of customer reviews for positive feedback, a relatively low risk application. The same restaurant could use the very same service to scan CVs for key words as part of shortlisting job applicants with certain skills, which could have fundamental rights implications if not subject to appropriate controls. [1]

In addition to consuming general-purpose AI models made available as services, customers are increasingly combining multiple AI models, often from multiple suppliers, to build more sophisticated AI systems. For example, a retailer might build an AI-powered customer service system that combines Microsoft's text analytics service to assess the urgency of a customer inquiry, with a pre-trained text generation technology from another supplier to auto-generate responses for certain customer inquiries. It could add another AI model from yet another supplier that matches each remaining inquiry to the most appropriate customer service representative. The ways in which different generally applicable AI technologies can be combined into AI systems, and the purposes for which customers can deploy these systems, are vast and varied. A system similar to the customer service system above could be developed and deployed in a higher risk setting, for example to triage and route patients calling a healthcare provider. Again, it is the customer that is combining these AI technologies into a final system for their specific purpose and adding their people and processes to that system. In doing so, the customer invariably takes the key decisions around deployment that can affect associated risks.

AI services can be customized with customer data and are dynamic

Customers are increasingly utilizing customizable AI services, such as Microsoft's Custom Voice and Custom Vision. This involves a customer taking an AI service that has been trained by the supplier to perform a general task—for instance, Microsoft's Custom Vision has been trained to recognize

[1] Note that Microsoft does not recommend the use of the Azure Cognitive Services Text Analytics API for recruitment purposes. We communicate the capabilities and limitations of our Text Analytics service via a Transparency Note that is accessible here.

objects—and conducting a second round of training on the model using the customer's own data in a process known as "fine-tuning".

For example, Minsur, a mining company committed to sustainable extraction, used Microsoft's Custom Vision to ensure that it could identify whenever any non-permitted effluent from its mine was entering a local waterway. They took Microsoft's Custom Vision AI service and "fine-tuned" the underlying model with their own labelled images from a camera taking pictures of the local waterway, training the system to identify water discoloration and other indicators of effluent in the river. The same Custom Vision model could be "fine-tuned" by a different customer for a different purpose, for example, to detect variation in tumor size as part of a radiology application. When "fine-tuning", the customer not only takes the key decisions around how and where to deploy the AI model, they also use their own data to alter the way the model performs. Depending on the nature of the data used for such fine-tuning and other factors, this too may introduce new risks.

In addition, many AI systems are dynamic, in the sense that their performance can adapt throughout their lifecycle as they learn from the data they process in operation. Using the preceding example, an AI vision model that is trained to detect variations in tumor size may become more accurate over time if designed to "learn" from the data (images of tumors) input into the system during its operation.

AI services are becoming more generally capable

Currently, most AI models that are broadly available on the market tend to perform one type of task (such as image processing) on one type of input (such as images). Research breakthroughs in 'multi-modal' models, however, suggest that increasingly a single AI model will be able to perform a range of different tasks based on different types of inputs, for example, being able to ingest video data and generate related text while also ingesting text to generate audio content. This increased range of functionality will further expand the range of scenarios in which entities can deploy AI technologies for different purposes. It will also reduce the ability of a supplier of the AI model to anticipate the full range of deployment scenarios and their associated risks, further reinforcing the need to focus in on the actual use of the AI system by the deployer.

Given these differences between AI systems and the physical products typically addressed by the NLF, Microsoft believes the AI Act should be adjusted to better address the nature of AI systems and the roles of different actors across the AI value chain. On our reading of the current proposal, the obligation-shifting mechanism in Article 28 creates significant legal uncertainty and produces confusing results, such as the scenario where a "user" of a system is routinely reclassified as the "provider." There are also currently very few requirements on "users" of general-purpose AI services, even though the sociotechnical risks AI systems present are heavily influenced by choices those actors make when they combine general-purpose AI services with their people and processes.

To better secure legal certainty and stability as essential preconditions to innovation, we recommend creating the designations of "technology supplier" and "deployer," in particular for the high-risk AI systems listed in Annex III. We believe that **both** of these changes are necessary to allocate the AI Act's regulatory responsibilities in a way that sees them fall on the entity best able to identify and effectively mitigate a system's associated risks. We detail these, and related, recommendations below:

Create a new role of "deployer"

Currently, the AI Act places most of the responsibility for complying with its obligations on "providers" of high-risk AI systems. However, as demonstrated above, in many cases it will be the entity that **deploys** an AI service that will determine both the intended use of the final AI system – including whether to use the system in one of the high-risk scenarios listed in Annex III—and the societal context in which the system operates. The deployer will also often be the only entity with full visibility into the system's operation and whether all relevant risks have been appropriately mitigated. Accordingly, it is appropriate that deployers of such AI systems take on responsibilities for addressing those risks that are within their control.

As such, we recommend creating a new designation of "deployer," defined as the entity that takes the specific decision to implement an AI system for one of the high-risk scenarios detailed in Annex III. We also recommend that this entity be responsible for ensuring that any such Annex III deployment satisfies the requirements set out in Article 16. This approach has the virtue of ensuring that regulatory responsibilities fall in the first instance on the entity that has the greatest control over, and visibility into, the operation of the specific deployment that brings it within scope of Annex III (and thus subject to the requirements of Articles 9-17). It is, however, contingent on "technology suppliers" also assuming responsibilities that they are well-placed to bear, as described below.).

Create a new role of "technology supplier"

For "deployers" to be effective in complying with their obligations under the AI Act, entities that supply AI technologies to these deployers will have important responsibilities to meet too. We note that Recital 60 already references technology suppliers indirectly, referring to "third parties, notably the ones involved in the sale and supply of software, software tools and components, pre-trained models and data." We urge policymakers to carry over this reference into the body of the Act by creating a new definition of "technology supplier" in Article 3.

In particular, the Act should make clear that "technology suppliers" would be expected to assist "deployers" (again, defined as the entity that makes the specific decision to deploy an AI system in one of the high-risk scenarios listed in Annex III) in complying with their obligations. This might include, for instance, providing information about the supplier's design choices when building the relevant AI model, including information about how the model has been trained and tested prior to release. Supplier-initiated testing could be against standardized benchmark tests that allow deployers to make 'apples-to-apples' comparisons against competing technologies made available by different suppliers. Technology suppliers could also provide documentation on the model's known limitations, including considerations that deployers should have in mind when choosing use cases. Ultimately, as outlined in Recital 60, technology suppliers should 'cooperate, as appropriate, with . . . [deployers] to enable their compliance'.

Microsoft currently provides information to customers of our AI services via our Transparency Notes. With a view to helping customers make informed deployment decisions, our Transparency Notes outline the capabilities and limitations of the AI technology we make available, as well as key considerations for how to use it responsibly. For example, in our Anomaly Detector Transparency Note, we outline how the service works, as well as a series of considerations for customers, including how to properly tailor the model with new training data, example use cases, and considerations when choosing a use case. We also outline scenarios in which Anomaly Detector is less appropriate for use, and the importance of human review of outputs when the system is being used in an environment where there may be a danger to physical safety.

The approach to regulation that we outline above will be familiar from other settings in which regulated actors rely on third-party suppliers to comply with their own regulatory obligations. For instance, financial services firms regularly rely on information and assurances from their technology suppliers to demonstrate that they satisfy their regulatory obligations, for instance around cybersecurity and resilience. Public authorities likewise often require technology suppliers to commit to complying with technical standards or other requirements as a condition of bidding on public procurement contracts. Although these regulatory obligations fall in the first instance on the financial firm or public authority at issue, these entities rely to a large degree on their technology suppliers (often via contract) to demonstrate compliance. We think a similar model would be more appropriate for the types of AI systems implicated by Annex III than the product safety framework that the Act currently imposes on such services.

2. Adjust the AI Act’s requirements to focus on outcomes and processes

The AI Act sets out an ambitious goal of establishing a single, uniform set of requirements that would apply across a broad and varied universe of AI products, services, and deployments. We applaud this ambition and believe this goal is achievable. We also think that the AI Act would benefit from a different approach in articulating these requirements.

Specifically, to be effective in securing its goals, the AI Act should recast the requirements set out in Articles 9 through 17 in terms of the **outcomes** they seek to achieve, together with a high-level description of the **steps** that regulated actors should adopt to achieve them. In other words, each requirement should state with clarity **what** it seeks to achieve, but give regulated actors greater leeway to determine, for their specific circumstances, **how** best to achieve it. This more flexible approach would help the AI Act more effectively address the wide range of different systems and use cases that it covers. It would also help the Act remain relevant in the face of ongoing technological developments and advances in the practice of responsible AI, including by allowing regulated actors to utilize evolving best practices and state-of-the-art tooling, both of which are important in enabling efficient and effective compliance.

Although we appreciate that certain requirements set out in the AI Act already follow this approach (for instance Article 13(1), which requires high-risk AI systems to be “sufficiently transparent to enable users to interpret the system’s output and use it appropriately”), others do not. For instance, several provisions set out requirements in prescriptive, formalistic terms that may be unworkable or impossible to achieve in some scenarios (we describe several examples below). Others are too specific—they zero in on particular types of mitigations, which might not be appropriate or effective for certain AI systems, while overlooking other mitigations that might be more suitable and effective. Also, many are not future-proof or innovation-friendly—they are based on a few examples of how AI systems are developed and used today, which could quickly become obsolete given the fast pace of AI innovation.

To illustrate this concern, consider Article 10, on data and data governance. A key goal of this Article is to ensure that high-risk AI systems operate in a fair and non-discriminatory way. As currently written, however, several of the obligations in Article 10 are infeasible, or are too narrowly drafted to achieve this goal in all scenarios. By way of example, Article 10(3) requires training, validation, and testing data set to be “relevant, representative, free of errors and complete.” Our experience through our internal responsible AI program leads us to believe that this requirement will be unworkable in some scenarios and inadequate in others. We explain why below.

Error-free data

Establishing appropriate processes to minimize the risk of errors in data capture, labeling, and other processing is a critical part of the data gathering and training process. However, in many contexts, it is difficult to define what “error-free” means. For example, language models have tremendous potential to help solve critical societal challenges in domains such as healthcare, where there are large amounts of unstructured text from which AI systems could help researchers identify patterns, trends, and connections. But data from these domains often reflect people’s beliefs, assumptions, attitudes, and values, as well as their ambiguity and uncertainty. Language is also very dynamic and contextual, again reinforcing the challenge of defining what "error-free" means. There are also situations where including errors in datasets can be advantageous. In the healthcare context, for example, the dataset used to train an AI model might include inaccuracies or other errors that cannot be cleaned due to their source (e.g., self-reporting of symptoms). But training an AI model on data that includes such inaccuracies might be helpful and even preferable where there is reason to expect that such errors might appear in input data once the system is put into service (e.g., in a hospital).

Complete and representative data

Assessing how well groups who will be subject to an AI system are represented in training datasets is critical to building a system that provides a similar quality of service to those groups. However, it is not possible to create a "complete" dataset even when working to ensure key groups are represented, because there is always more data that can be collected. Among other reasons, this is because the world in which we live is ever-changing and so too is its data trail. Incomplete datasets can also be helpful to overall system performance in certain circumstances, for example, where incomplete records provide unique information that is valuable to the operation of the system.

Moreover, ensuring that training, testing, and validation datasets are appropriately tailored to the task at hand is only part of the solution to promoting AI fairness and non-discrimination. Regulated actors also need to make careful choices about model functionality and design. They also need to test systems, including in the real-life settings in which the AI system will operate, because results achieved "in the lab" often cannot capture the full range of scenarios that may occur once the system is deployed. This real-world testing of actual deployments needs to occur on an ongoing basis in order to identify issues that may only emerge over time. Because only deployers of AI systems will be able to do this real-world testing in many scenarios, it isn't clear that Article 10 — or the AI Act as a whole — would provide the level of fairness and non-discrimination in AI systems to which it aspires.

In our view, an outcomes-based approach to requirements is more likely to achieve the AI Act's goals. This approach would begin by clearly articulating the outcomes that regulated actors should strive to achieve in order to promote EU values of fairness and non-discrimination in the context of high-risk AI systems. By way of illustration, such a requirement could provide that:

- High-risk AI systems should provide a similar quality of service[2] for relevant demographic groups impacted by the system.
- High-risk AI systems that allocate resources or opportunities[3] should do so in a manner that minimizes disparities in outcomes for relevant demographic groups impacted by the system.
- High-risk AI systems that describe, depict, or otherwise represent people, cultures, or society should minimize the potential for stereotyping, demeaning, or erasing relevant demographic groups impacted by the system.[4]

These goals could be supported with clear steps that regulated actors would take to help meet these goals, for instance:

[2] Quality of service refers to whether an AI system works as well for one person as it does for another, even if no opportunities, resources, or information are extended or withheld.

[3] Such as in finance, education, employment, healthcare, housing, insurance, or social welfare.

[4] In practice, approaches to measure and mitigate quality of service harms are much better understood than how to measure and mitigate representational harms (i.e., harms that involve stereotyping, demeaning, or erasing relevant demographic groups) and the AI Act needs to reflect this reality.

- Identifying the relevant demographic groups impacted by the system;
- Designing and undertaking an evaluation to assess the extent to which the goal is achieved;
- Reassessing the system design (including the training data, model features, objective functions, training algorithms, and approaches to human-AI interaction) to mitigate gaps revealed by the evaluation;
- Re-evaluating the system after incorporating appropriate mitigations; and
- Communicating material residual risks to deployers so that appropriate precautions can be taken, including decisions to not use certain systems in particular use cases if they are not fit for purpose.

In each case, achievement of the desired outcomes and execution of the procedural steps would need to be assessed by reference to the state-of-the-art and industry best practices, along with a clear-eyed recognition of the highly contextual nature of fairness and the fact that it is never possible to fully "de-bias" an AI system or "guarantee" its fairness.

Rather than imposing specific measures or mitigations on regulated actors—as the AI Act currently does in many cases—this outcomes-based approach would give regulated actors the ability to adopt whichever set of measures or mitigations were best suited to achieving the enumerated goals within the context of the specific technology, scenario, and deployment at issue. It will also spur further development of measurement and mitigation techniques, and the preparation of harmonized standards or common specifications that can be tailored for specific scenarios, for example, acceptable false negative rates for systems used for clinical diagnoses of cancer. We urge policymakers to review Articles 9 through 17 of the AI Act with this in mind, and to revise those provisions to ensure that they are sufficiently outcomes-oriented to be appropriate and effective across the full range of AI systems to which the AI Act applies.

We set out below several specific changes that in our view would help align the Act's substantive requirements with the approaches set out above. At same time, we urge policymakers to consider these changes irrespective of the approach finally chosen.

Article 9: Set out the types of risks that must be assessed by risk management systems

Although the recitals state that the AI Act seeks to address risks to health, safety, and fundamental rights,[5] Article 9 does not include similar language, nor does the AI Act define the meaning of the term "risk." Article 9 would achieve greater legal certainty if the text of Article 9 itself specified the types of risks that must be addressed. Greater clarity on expectations for assessing rights-based risks in Annex III would be particularly helpful.

Article 14: Ensure human oversight obligations are feasible

Currently, the AI Act addresses the issue of effective human oversight by placing design requirements on the "provider" of an AI system with a view to minimizing risks to "health, safety and fundamental rights." This is an important goal, but one that cannot be fully realized by system design alone. As such, we make the following recommendations:

[5] See Recitals 1, 13, 27, and 43.

Provide further information around the definition of human oversight and the goals of Article 14.

The AI Act would benefit from further defining what it means by "effective human oversight" and the specific outcomes this section seeks to advance. A human cannot be assumed to oversee every AI action. What constitutes "effective" human oversight will also differ markedly depending on the deployment scenario and the nature of related risks. As such, we recommend that the AI Act instead require deployers to implement sufficient, qualified human oversight as is appropriate to the deployment scenario at issue.

Ensure deployers provide adequate staffing and training.

Article 14 requires providers to ensure systems are designed in a way that enables effective human oversight. Making such oversight meaningful, however, will also require deployers to ensure that the humans performing the oversight are trained and equipped appropriately. Careful consideration should be given to an overseer's level of skill and training, consistent with the instructions of use and other information provided by the supplier, and to whether the overseer's exercise of oversight is tied to the intended use of an AI system. It will also be important to create accountability mechanisms to assess the effectiveness of the human overseer.

Limit two-person oversight requirement in Article 14(5) to law enforcement or similar uses.

Requiring two-person review of biometric identification systems for all uses is disproportionate and counterproductive in many cases. For instance, it would make no sense to require two-person verification when such a system is used in a factory as a safety measure to ensure that only qualified workers operate certain dangerous machinery, or in a retail store to ensure that only authorized employees have access to cash registers. To remedy the overbreadth of Article 14(5), it should be limited to specified, uniquely high-risk scenarios, for instance where the misidentification of a person could have legal or similarly significant effects on an individual. In addition, Article 14(5) should require providers only to **enable** (not "ensure") two-person oversight.

Article 17: Allow any type of management system

The management system required by Article 17 should be inclusive of the types of generic "management systems" that are regularly applied to digital services. It should not be limited to "quality management systems," which are primarily used by traditional manufacturing industries in respect of manufactured goods, given the unique features of many AI services highlighted in section 1 above. This more inclusive approach to what constitutes a qualifying "management system" also has the benefit of allowing many regulated actors to rely on their existing control frameworks and thereby minimize their compliance costs – a key concern for start-ups and small and medium enterprises in the EU's AI ecosystem.

3. Adapt the post-market monitoring regime to better address the sociotechnical nature of many AI risks

To ensure the AI Act succeeds in creating greater legal certainty around AI development and deployment in the EU, we offer the following recommendations with a view to securing a viable and stable post-market monitoring and enforcement regime that aligns with the sociotechnical nature of AI systems

Article 21: Require immediate corrective action only where feasible and appropriate

Requiring regulated actors to take "immediate" corrective action when a system is not conforming with the AI Act will be infeasible and counterproductive in many scenarios. It can take time to correctly identify, assess, and rectify certain issues, bearing in mind that sometimes these steps require the AI technology supplier and/or deployer to conduct analysis, testing, and the collection of new data. To address this, we would recommend revising Article 21 to clarify that corrective action must be reasonably prompt under the circumstances, taking into account any analysis and testing needed to remedy the issue.

Article 43: Refine the obligation for a "substantial modification" to trigger a new conformity assessment

We appreciate the intent of Article 43(4) — namely, to ensure that high-risk AI systems that undergo significant changes after being placed on the market do not present new or greater risks than the original system. It will be important to clarify the definition of "substantial modification" to ensure that only material adverse changes are captured, as well as to reflect the decisive role deployers play in modifying a system. If the criteria for a "substantial modification" are set too low, there will be a direct correlation in prohibitively expensive re-certification costs and a potential for suspension of services that could materially harm European industry from being globally competitive.

Article 61: Share post-market monitoring obligations across the value chain

For most systems used in a high-risk scenario—in particular those listed in Annex III—deployers will hold or control the relevant input data, and normally will be in the best position to monitor the system's operation and quickly identify any malfunctions or other problems. Technology suppliers, by contrast, may not have the ability to do any of these things—especially where the AI system is used in a sensitive or highly regulated environment (e.g., a bank, hospital, law enforcement agency, etc.). In light of this, and consistent with our comments in section 1 above, we urge lawmakers to impose the post-market monitoring obligations on the entity that is closest to the system's operation. In the case of Annex III systems this will generally be the deployer. We also urge lawmakers to allow the deployer to contractually allocate responsibilities to all parties involved in the AI system (including technology suppliers), as appropriate to each party's role in the supply chain. This is a practice followed in the medical devices industry where the regulated entity bringing a registered medical device to market defines obligations via a "Quality Agreement" with technology suppliers in a way that allows the regulated entity to meet its compliance obligations.

Article 64: Add safeguards to market surveillance authorities' right of access to data and source code

Article 64(1) authorizes market surveillance authorities to demand "full access to the training, validation and testing datasets used by the provider" (Art. 64(1)), and to the AI system's source code upon a "reasoned request" (Art. 64(2)). Given the security, privacy, and trade secret risks associated with sharing source code and datasets, and the availability of effective auditing methods that do not require such steps, we believe that market surveillance authorities should be required to first use those other methods to assess AI system functioning and compliance, in line with the proportionality principle. Should authorities still consider access to datasets and/or source code essential after applying these methods, we think judicial authorization and oversight is an appropriate and necessary safeguard.

4. Further changes to promote innovation and strengthen fundamental rights protections

We welcome the Commission's efforts to craft the AI Act in a way that provides space for research and innovation while also protecting fundamental rights, and we particularly support the Act's provisions to protect rights in connection with law enforcement uses of real-time remote biometric identification systems in publicly accessible spaces. To help advance these goals, we offer the following further suggestions:

Article 3: Exclude the use and distribution of AI systems for research and development purposes from the scope of the AI Act

We suggest the terms "placing / making available on the market" and "putting into service" should specifically exclude use of AI systems for internal research and development (R&D) purposes and distribution of AI systems for R&D generally. AI researchers and software developers regularly upload AI models and other AI-related materials to repositories such as GitHub (which is owned by Microsoft). Open-source repositories play a critical and highly beneficial role in the software ecosystem. As Article 3 is currently written, there is a risk that those who upload these materials to software repositories, or the operators of these repositories themselves, could be viewed as a regulated entity, without ever having taken any active step to "put into service" the system in the EU. This might inadvertently undermine AI research and open-source software innovation in the EU. We therefore recommend redrafting Article 3 to exclude those activities.

Article 5: Further strengthen the AI Act's fundamental rights protections for certain law enforcement uses of remote biometric identification technology

Microsoft supports the way in which Article 5 addresses the issue of law enforcement use of real-time remote biometric identification systems in publicly accessible spaces. We believe that these technologies can provide significant societal benefit in helping increase public safety and security. However, we also recognize the significant risks they may pose to fundamental rights if appropriate guardrails are not enacted. As such, we believe that such systems should be used by enforcement in these scenarios only when strictly necessary and in a targeted and responsible manner. Microsoft believes that the Article 5 provisions can be further strengthened by:

- **Further limiting the types of crimes for which the systems can be used**

  In addition to the targeted search of victims and imminent threats specified in Article 5(1)(d)(i) and (ii), Article 5(1)(d)(iii) permits the use of real-time remote biometric identification systems by law enforcement for the investigation of crimes referred to in Article 2(2) of Council Framework Decision 2002/584/JHA that are punishable in the Member State concerned by a custodial sentence of a maximum period of at least three years. That list of crimes, however, includes a range of lower level, non-violent crimes, including forgery and counterfeiting, for which use of real-time remote biometric identification systems in publicly accessible spaces seems to be disproportionate. We recommend further discussion on this issue and potentially further limiting law enforcement use of such systems to the most serious and violent crimes, for example murder, armed robbery, and kidnapping.

- **Requiring each individual use of the technology to be subject to prior judicial authorization**

  The Act currently requires each use of real-time remote biometric identification be subject to prior authorization via judicial authority or independent administrative authority of a Member State. We strongly support this need for authorization which is important in ensuring public oversight and democratic accountability around law enforcement use of these technologies.

We recommend narrowing this section to permit judicial authorization only. This will serve as an important protection for fundamental rights, combined with the Article's requirement to balance the benefits of using the system in a permitted scenario with consideration of the impact of system use on rights and freedoms.

- **Providing for greater transparency, accountability, and public scrutiny**

  Transparency and accountability around how law enforcement uses technology is essential to building and maintaining public trust and ensuring democratic scrutiny. We recommend that the Act require:

  - **Accountability reporting**

    Law enforcement agencies using real-time remote biometric identification systems in publicly accessible spaces should create a publicly available accountability report that includes:

    - A description of the technology and how it works, including capabilities and limitations;
    - The specific purposes for which the technology will be used;
    - Data handling procedures;
    - Testing procedures, including a process to periodically undertake independent tests of the performance of the system in operational conditions; and
    - Training and personnel procedures, including ensuring adequate training and that all decisions are subject to human review.

    We also recommend that any government agency using real-time remote biometric identification systems in publicly accessible spaces be required to prepare an annual report that discloses the extent of their use of these systems.

  - **Transparency reporting of judicial authorizations**

    We recommend establishing a system for transparency reporting of the judicial authorizations given. Reporting could detail: 1) general information on the authorizations sought and granted; 2) the number and duration of each authorization; and 3) a description of the nature of the public spaces where real-time remote biometric identification was conducted.

- **Creating additional fundamental rights protections**

  We recommend adding further protections to help ensure that these technologies are not used in a way that undermines fundamental rights. These should include:

  - An explicit prohibition on law enforcement uses of real-time remote biometric identification in publicly accessible spaces to record any individual's exercise of freedom of assembly, expression and association; and
  - An explicit prohibition on law enforcement uses of these technologies based on an individual's religious, political or social views or activities, participation in a particular non-criminal organization or lawful event, or actual or perceived race, ethnicity, citizenship, place of origin, age, disability, gender, gender identity, sexual orientation or other characteristic protected by law.

Article 51: Require registrations of deployments, not systems

We appreciate the apparent goal of Article 51 — to ensure that people whose health, safety, or fundamental rights are potentially at risk from a high-risk AI system are aware of those risks, can engage in a debate about the relative benefits and risks of the system, and can, if they wish, take steps to avoid the system.

In our view, that goal would be better secured by a different approach. From a consumer protection and fundamental rights perspective, there is relatively little value in knowing that company X has offered high-risk AI system Y on the EU market. Of much greater value to consumers and citizens would be information on whether any such AI system has been deployed in a location or situation that is likely to affect them—e.g., in a school their child attends, in a bank where they are a customer, in their local hospital, or by a relevant government agency. This point is likely to be particularly true with regard to the high-risk AI systems listed in Annex III because, with respect to those systems (as compared to the products listed in Annex II), there is a significantly greater chance that: (1) people who are potentially affected by the system might otherwise not know that they might be so affected; and (2) the types and degrees of risk are likely to vary significantly depending on the specific deployment of the system and the surrounding context and facts of deployment.

To address this concern, policymakers should consider modifying Article 51 to require registration of deployments of high-risk AI systems in the environment where they are in operation,[6] at least with respect to the AI systems listed in Annex III.

Conclusion

Microsoft thanks the Commission for this opportunity to share our initial comments on the proposed AI Act. We recognize the scale of the task that the Commission has undertaken and look forward to sharing any additional thoughts as we continue to study the proposal and engage in dialogue about it. Microsoft is committed to playing its part to help the EU embrace AI technologies safely and in ways that respect fundamental rights and European values. We look forward to engaging with EU policymakers, our European partners and customers, and other stakeholders to support the development of the Act.

[6] See, for example, the City of Amsterdam Algorithm Register (last accessed on 1 August 2021).