TRUST AND EXCELLENCE MADE IN EUROPE

Response of Microsoft Corporation
to the European Commission's White Paper
on Artificial Intelligence

Microsoft

June 2020

Introduction and Executive Summary

Microsoft appreciates the opportunity to offer these comments on the European Commission's White Paper on Artificial Intelligence — A European Approach to Excellence and Trust ("White Paper"). We support the goal of promoting the development and uptake of artificial intelligence ("AI") across the EU in a way that respects European values. We commend the Commission for its initiative in proposing a groundbreaking regulatory framework for high-risk AI, and for its bold proposals to make Europe a world leader in AI.

AI offers tremendous opportunities for people and societies across Europe. Every day, we see people using AI to create innovative new products and services, make existing offerings better and safer, tackle major societal challenges, or simply make life more enjoyable. By giving computers the ability to do tasks once reserved for humans—derive insights from data, draw inferences, and even see, hear, and comprehend—AI solutions are enhancing human capabilities, leading to better outcomes and improved lives. The new realities wrought by the COVID-19 pandemic, and the many innovative responses to these challenges across the EU, further highlight the positive and empowering potential of AI.

Recent developments also demonstrate the importance of addressing social inequities. While AI can be part of a solution to social challenges, we also need to make sure it is not part of the problem. The ongoing calls for justice and equality, and for the end to social disparities, demand that we examine how we use AI and what we need to do collectively to ensure AI is developed and deployed in a manner that promotes our shared social goals.

> **"AI isn't just another piece of technology. It could be one of the world's most fundamental pieces of technology the human race has ever created."**
>
> Satya Nadella, CEO, Microsoft

Policymakers and regulators play a critical role in promoting an ecosystem of trustworthy AI through their actions. The unique power of AI, and its application across so many sectors, domains, and use cases, represents a paradigm shift in computing, one that is only beginning to unfold. Ensuring that AI is trustworthy—that it is safe, ethical, and accountable—will require a corresponding cultural shift, both by those who develop and supply AI technologies to the market, and those who deploy and use it in real-world settings. Microsoft believes that companies that create technology also have a responsibility to help secure the promise of its future

In these very early days, we view trustworthy AI as a journey more than a destination. It is one we care about deeply, and we welcome the opportunity to work with the Commission and other stakeholders to advance towards a society in which trustworthy AI is the norm. In some settings, this might require new laws, or new interpretations of existing laws. The challenge is to do so in ways that do not impede the vast positive potential of AI.

We thus offer these comments not because we oppose AI regulation, but rather to aid the Commission in the difficult task of assessing where regulation might be most appropriate and how best to regulate consistent with European values. As elaborated in Part I, our principal suggestions on the EU's proposed AI regulatory framework are as follows:

Promote trustworthy AI through governance and tools.
Regulatory frameworks for AI should incentivize relevant actors to adopt governance standards and procedures that support their efforts to operationalize trustworthy AI, and should support the development of technologies, systems, and tools to help these actors identify and mitigate relevant risks.

Leave space for positive uses of AI.
The use of AI can help make some products and services safer and better than their non-AI counterparts. Policymakers should take care to ensure that the cost of AI regulation is not so high that it prevents these products from reaching the market.

Differentiate types of harm.
Risks to safety and risks to fundamental rights are inherently distinct; any AI regulatory regime should recognize this distinction, both in the requirements it imposes and the compliance regime it adopts. Both are important to address.

Clarify roles of regulated actors.
AI regulation should be clear on which requirements fall on which regulated actors (developers, deployers, etc.) and should impose responsibilities on the actor that can most efficiently and effectively comply with them.

Leverage existing laws and regulatory frameworks.
Absent clear gaps, policymakers should rely on existing laws to the extent possible rather than adopt wholly new regulatory frameworks and obligations on top of them. Where new laws are needed then they should be adopted.

With regard to the specific proposals set out in the White paper, we would encourage the Commission to:

- Adopt a clear definition of AI;
- Adopt a differentiated and holistic conception of risk;
- Refine the definition of "high risk" AI;
- Reconsider the application of the proposed regulatory requirements;
- Clarify the addressees of the regulatory obligations;
- Adopt light-touch compliance and enforcement mechanisms;
- Adopt a flexible voluntary labeling regime; and
- Further assess whether changes to the EU's safety and liability frameworks is necessary.

Part II responds to the Commission's proposals for an AI ecosystem of excellence and discusses various actions that Microsoft is taking in this area. Part III describes Microsoft's journey to responsible AI, and Part IV highlights several resources we have developed to help put responsible AI principles into practice.

AN ECOSYSTEM OF TRUST

We welcome the Commission's AI initiative. The EU has the opportunity to design a regulatory framework for AI that could advance the responsible development and use of AI not just in Europe, but globally. It is important that as the Commission seeks to ensure that AI in Europe evolves in a manner that respects its values and fundamental rights—a goal we fully support. Equally important, any framework should be interoperable with the many other ongoing efforts to promote trustworthy AI in other jurisdictions and forums.

We commend the Commission for proposing an incremental, risk-based approach to regulation. An incremental approach—one that imposes mandatory requirements only on a discrete and clearly defined set of "high-risk" scenarios—will enable the technology to advance and at the same time restrict harmful uses.

We set out our reactions to specific aspects of the Commission's regulatory proposal below.

Overarching observations

We offer the following broader, thematic points, which we hope may inform the Commission's actions as it considers how best to move forward on AI regulation:

Promote trustworthy AI through governance and tools.

Over the past 24 months, there have been significant advances in the global discussion on AI, with many organizations and governments articulating foundational principles for trustworthy AI. The challenge for all stakeholders is now how to operationalize these principles across the many different types of AI technologies that exist, and the almost infinite variety of domains and scenarios in which these technologies will be deployed.

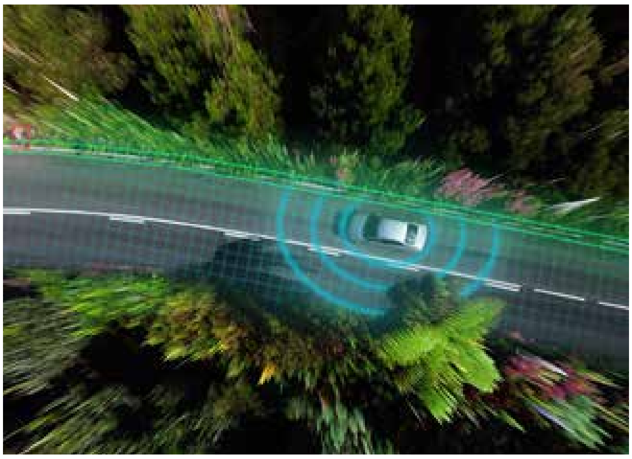
We should not underestimate the magnitude of this challenge. Based on our own journey in responsible AI, we think it is imperative that regulatory frameworks for AI focus first and foremost on two key objectives: first, to incentivize relevant actors to adopt governance standards and procedures that will support their efforts to operationalize trustworthy AI; second, to support the development of technologies, systems, and tools to help these actors identify and mitigate relevant risks.

The first goal is critical to translating trustworthy AI principles into practice. The second goal recognizes that implementing trustworthy AI in the real world is not only technically difficult, but can also raise exceedingly complex social and ethical questions as well, such as how to balance an AI system's clear benefits to one group against potential risks to another, or how to assess what is and isn't "fair." The answers to these questions may vary depending on the AI system at issue, how it is deployed, and the purposes for which it is used. As described in Annex II, Microsoft is working hard to create such tools because we have learned that responsible AI is a much more complex task than simply telling product teams to "use representative datasets" or "achieve fair outcomes." Having tools that can help people identify risks, and devise mitigation strategies from the design stage onwards, is not simply a "nice to have"—it is an imperative, necessary and essential to trustworthy AI.

Incentivizing organizations to adopt robust internal governance, and equipping them with tools to identify and mitigate risk, is in our view more likely to be effective at advancing the goals of trustworthy AI and be more scalable than a regulatory regime that mandates specific outcomes. Indeed, given the almost infinite variety of AI systems, and the scenarios and domains in which they can be deployed, we believe that simply requiring AI systems to have certain characteristics or achieve certain outcomes is unlikely to work. This is not to say that outcomes, like mitigating bias should not be considered.

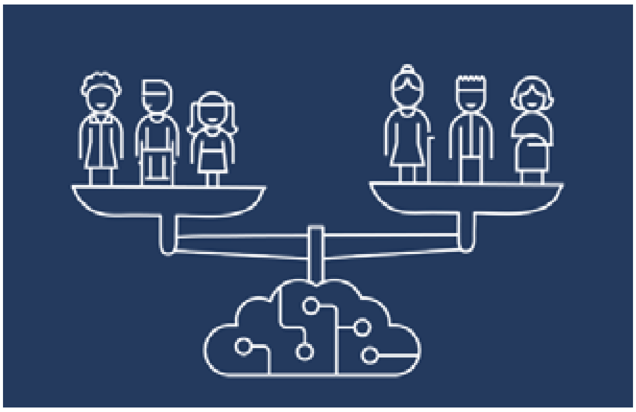
Leave space for positive uses of AI

In considering how best to address the risks AI may raise, it is important to bear in mind that, in many cases, the use of AI—even in high-risk scenarios—may actually make products and services safer, better, or more accurate than their non-AI counterparts. Consider, for example, autonomous vehicles. There is broad recognition that the adoption of self-driving vehicles is likely to make road transportation significantly safer than it is today. Similarly, researchers across the world are finding new ways to use AI to make medical devices and tests safer, more effective, and more accurate. Indeed, one of the primary drivers of AI adoption in many scenarios is that it can make existing products and services, or the environment in which they are built, better and safer. This is not to deny that autonomous vehicles or AI-powered medical devices (or any other AI system) might also raise unique risks, or to say that such risks should not be regulated. But when weighing how to regulate AI systems, care should be taken to ensure that the cost of regulation does not become so high that it prevents safer and better products and services from reaching the market.



Differentiate types of harm.

Discussions on the risks AI systems may pose generally highlight two types of harms: risks to safety (e.g., physical harm) and risks to fundamental rights. These are two inherently different types of harm, and it is vital that any AI regulatory framework recognizes this distinction, both in the requirements it imposes and the compliance regime it adopts. Consider, for example, the requirement that training data be adequately representative of affected populations. While that requirement might be quite relevant to mitigating risks to fundamental rights—e.g., in a facial recognition system—it is likely to be irrelevant to mitigating the safety concerns raised, for instance, by autonomous construction machinery. Conversely, pre-marketing conformity assessment might be an appropriate way to ensure that autonomous construction machinery is safe, but quite ill-suited to evaluating the fundamental rights risks in an AI voice-recognition service that is continually improving and available for implementation in thousands of different settings. This distinction between risks to safety and fundamental rights is already deeply embedded in the EU regulatory acquis, with product safety regulated largely on a sectoral basis and fundamental rights legislation being primarily horizontal in nature. Similarly, any requirements on AI systems, and the corresponding compliance and enforcement regime, should clearly identify the type of harm that it seeks to address and be appropriately tailored to such harms.



Clarify roles of regulated actors.

Different actors in the AI ecosystem have different roles to play in promoting trustworthy AI. Developers, for instance, often will be best placed to identify and mitigate certain risks (e.g., those that can be clearly identified and efficiently addressed in the design or development phase of an AI system), while deployers likely will be best suited to address others (e.g., those that vary depending on the scenario in which the system is used and the affected populations). Any legislative proposal should be as clear and precise on this allocation of roles and responsibilities as possible. To be effective, AI regulation should leave no uncertainty about which requirements apply to which actors and should ensure that responsibilities always fall on the actor that can most efficiently and effectively comply with them. Similarly, the compliance and enforcement regimes corresponding to these requirements should be appropriate to this allocation of responsibilities and should not require one set of actors to ensure compliance with requirements that reasonably should fall to another.

Leverage existing laws and regulatory frameworks.

The EU already has an extensive regulatory framework covering safety, security, consumer protection, fundamental rights, and other important interests. These laws apply equally to AI-powered products and services. Indeed, some of these laws specifically take into account risks that may be associated with AI (e.g., the General Data Protection Regulation's ("GDPR") rules on automated individual decision-making and processing of "special category" data).

There may be certain types of AI systems and applications that raise risks that are so truly unique that existing EU laws might not address them. As discussed in more detail below, Microsoft believes that the use of facial recognition systems by the public sector for consequential uses—and in particular their use in public spaces — is one example, but others may exist as well. Absent clear evidence of gaps in the current EU regulatory acquis, however, policymakers should rely on existing laws to the extent possible — and augment them with guidelines or other refinements where appropriate — rather than to adopt wholly new regulatory frameworks and obligations on top of them. Duplicative and overlapping requirements will add complexity and impose regulatory burdens without any clear corresponding benefit for individuals or society. At the same time, where new requirements are needed they should be adopted.

Comments on the proposal

The preceding comments apply broadly to all proposals for AI regulatory frameworks. The comments that follow address discrete aspects of the Commission's regulatory proposal set out in the White Paper.

Adopt a clear definition of AI

The term "artificial intelligence" can be, and often is, used to describe a vast array of technologies, including methods that perform computer-based perception, learning, and reasoning. These technologies can be used separately or combined to yield systems that perceive, classify, predict, or otherwise reason in an automated or autonomous manner, and can be used in a nearly infinite range of scenarios.

We agree that any future regulatory framework should apply to both the development and deployment of systems that use AI. But what is "AI" for purposes of the Commission's proposal? Consistent with the Commission's goal of carefully focusing regulatory mandates on high-risk scenarios, it will be crucial to define what technologies constitute "AI" for the purpose of this framework. Doing so is imperative to enabling regulated actors to understand whether their products and services fall within scope. Without a clear definition, the application and enforcement of any regulation will be both difficult and ineffective.

Adopt a differentiated and holistic conception of risk

Microsoft strongly supports the Commission's conclusion that any AI regulatory framework "should follow a risk-based approach" (p. 17). As the White Paper notes, such an approach "is important to help ensure that the regulatory intervention is appropriate" (id.). Although the White Paper identifies many different kinds of risks that might arise with regard to AI (pp. 10-16), it does not articulate a clear definition of "risk." Because the concept of risk is so central to the Commission's approach, and in some sense forms the policy foundation—the "why we are regulating"—to the proposed framework, we would encourage the Commission to ensure that any final legislative proposal explicitly defines the concept of risk. Further, this definition must be sufficiently holistic to capture the full range of considerations that arise in the context of use of AI.

To be effective as a policymaking tool, we believe this concept of risk must include two elements. First, it should expressly consider both the severity of potential harm and the likelihood that this harm will occur. For instance, an AI system that recommends songs based on a user's prior listening habits may appropriately be considered low risk irrespective of its quality because the potential severity of harm from a poor system (one that frequently recommends songs that users don't like) is so low. Conversely, an automated factory machine that has the potential to cause severe physical harm might nonetheless be considered low risk if it is only used in scenarios where the likelihood of injury is exceedingly low (e.g., it operates only when no people are in the vicinity). Especially if the final regulation requires regulated actors to identify or mitigate the risks of an AI system, it is vital that this concept of risk expressly takes into account both the severity and likelihood of the harm(s) at issue.

Second, the concept of risk also should take into account both the benefits of the AI system (for individuals and for society more broadly), and the "harms" of not adopting the AI system. By way of example, any analysis of whether and how to regulate the safety risks of autonomous vehicles to drivers and passengers should also take into account the potential safety benefits to others (e.g., pedestrians, drivers and passengers in other vehicles) and also the change in overall safety as compared to human-directed vehicles.

This does not mean that regulation is necessarily inappropriate when the overall benefits of an AI system outweigh the risks. But to the extent any regulation requires regulated actors to identify or mitigate risks, it should also require them to take account of these benefits as well.

Refine the definition of "high-risk" AI

The White Paper proposes that mandatory regulatory requirements would apply only to AI applications categorized as "high risk" (p. 17). An AI application would be deemed high risk only if it met two cumulative criteria:

1. It was "employed in a sector where . . . significant risks can be expected to occur," with these sectors being "exhaustively listed" in any new regulatory framework; and

2. It was "used in such a manner that significant risks are likely to arise," and proposes by way of example "AI applications that produce legal or similarly significant effects for the rights of an individual or a company; that pose risk of injury, death or significant material or immaterial damage; [or] that produce effects that cannot reasonably be avoided by individuals or legal entities" (id., emphasis added).

The White Paper adds that certain AI applications might be deemed high risk even when used outside a high-risk sector—for instance, where AI is used for "remote biometric identification" or "for recruitment processes [and] situations impacting workers' rights" (p. 18).

We support the Commission's plan not to impose new regulatory requirements on AI applications that do not pose a high risk of harm. As noted, the EU already has an extensive regulatory framework to protect consumers from a wide array of harms, including risks to safety, security, consumer rights, data protection, discrimination, and many others. We agree with the White Paper that these laws should apply to AI systems just as they do to traditional products and services. We also acknowledge that the application of these laws may require new interpretation due to the specific attributes of AI to ensure protections are in place.

The definition of high risk set out in the White Paper raises important questions and could be further refined. We offer the following observations and suggestions:

High-risk sectors

As the White Paper notes, it will be important for any final regulation to exhaustively list the covered sectors so that regulated actors can know whether they fall within scope. It is noteworthy that several of the sectors listed in the White Paper—healthcare, transport, and energy—are already subject to extensive sectoral safety and related regulation. For instance, the Medical Devices Regulation already covers medical devices that incorporate software, including AI-powered medical devices. Rather than create a new horizontal AI regulation that applies to them on top of these existing rules (for similar categories of risks), we urge the Commission to consider whether it would be more efficient and effective to extend those existing rules to cover new types of risks that AI might present in relation to the same safety risks.

One possible exception to this statement on existing sectoral rules is the White Paper's reference to the public sector. We agree that uses of AI by the public sector may raise unique concerns, given that many public-sector decisions can have consequential impacts on individuals' rights and freedoms and their ability to access essential benefits and services, among others. Given the many complex challenges of AI regulation, the Commission might consider focusing any horizontally applicable AI regulatory requirements in the first instance on public-sector uses of AI. This will enable the Commission to learn what works and what does not, and then to assess down the road whether to extend these rules to other sectors. Such an approach would indirectly regulate many private-sector entities as well, given that private companies are often the ones who develop and supply AI technologies used by the public sector.

High-risk uses

The White Paper's proposed definition of what constitutes a high-risk "use" is potentially quite broad. For instance, if a website uses AI to rank online ads, and as a result certain consumers do not see an ad offering low-interest loans, does this constitute a "legal or similarly significant effect"? Even if not, is the effect one that consumers "cannot reasonably . . . avoid"? Microsoft recognizes the importance of thinking broadly about how AI systems may impact people—indeed, this element of the proposed high risk definition is similar to one that Microsoft has proposed for developers to use in envisioning potential risks. However, to the extent this standard is one that regulated actors must interpret and apply in determining whether they are subject to mandatory regulatory requirements, it will be important for this definition to be precise, easy to apply in practice, and not overly broad in scope.

At the same time, where use of an AI application may present a high risk, the Commission's definition means it will only be subject to regulation if it is used in a high-risk sector as well. Developers of AI systems might not always know, or be able to control, the sectors in which customers use such systems, however. Further, the Commission's approach necessarily means that certain uses of AI that threaten a significant risk of harm may be excluded from regulation (i.e. to the extent they do not fall into a covered sector). It will be important for the Commission to explain how a compliance regime will operate in such situations or, at some point, to consider decoupling the sector and high-risk use requirements. We recommend keeping this aspect of the regime under ongoing review, to ensure that the Commission's regulatory approach does not leave meaningful risks unaddressed.

Finally, the White Paper proposes a further debate on the specific requirements applicable to remote biometric identification in public spaces. Microsoft welcomes this debate. We have long advocated for regulation of the use of facial recognition systems by the public sector, an exemplary high-risk use scenario that we strongly believe requires specific rules. Government use of such systems to monitor and specifically identify individuals raises significant risks not only to privacy, but also to other important fundamental rights such as freedom of expression and assembly. Bias in such systems can also meaningfully increase the risk of decisions, outcomes, and experiences that are discriminatory, exacerbating existing cultural and social biases.

The proposed regulatory requirements

The Commission's White Paper proposes various requirements that would apply to high-risk AI applications. These relate to training data, record-keeping, robustness and accuracy, and transparency, among other measures.

Collectively, the proposed requirements seek to promote both greater transparency and greater accountability in the development and deployment of AI. These objectives are broadly consistent with the HLEG's AI ethics guidelines and other AI ethical principles such as the OECD's principles for ethical AI and Microsoft's Responsible AI principles. Answering the Commission's question as to whether these are the "right" requirements is more difficult, however, because the answer depends on the context. Ultimately, regulatory requirements must map to the risks they seek to mitigate. Without clarity on the specific harms that each of these regulatory requirements is meant to address—and whether they will only be applied to AI systems that actually raise a high risk of such harms—it is difficult to assess whether the proposed requirements are fit for purpose. In order to develop requirements that meaningfully address the risks, we need to start with a more thorough mapping of requirements to risks, and risks to AI applications.

While certainly high-level goals, it is also not clear that imposing substantive or outcomes-based compliance requirements—for example, "AI systems must be accurate" or "training data must be sufficiently representative"—is the optimal way to regulate here. As a baseline, several requirements the Commission proposes could be useful to mitigate certain risks in certain scenarios. But it will often be the case that not all are relevant to any specific AI application, even if it is high risk. In other scenarios, it may be that none of the White Paper's proposed requirements is sufficient to mitigate the risks, and that different measures are needed. Of course, in other contexts, such requirements, e.g. accuracy, are critically important.

How a given risk is best addressed will be highly context-specific. Rather than imposing specific requirements for how AI systems are developed and deployed, requiring (or, in the case of non-high-risk AI, incentivizing) developers and deployers to have governance standards and procedures in place that ensure that they appropriately consider the harms that may arise from use of their technologies—and, where needed, to take steps to mitigate those harms—is likely be more effective.

For example, developers of high-risk AI systems might be required to adopt internal policies and procedures to promote the development of trustworthy AI.

These policies and procedures might include requirements for:

- product teams to envision the full range of harms that any new AI system might impose on individuals and society, and take appropriate mitigation steps;
- training those involved in the design, development, testing, and marketing of the AI system on these processes, and assign specific individuals or groups within the company with responsibility for overseeing implementation and compliance;
- transparency with customers, users, and other affected stakeholders about risks inherent in the use of the relevant AI system; and
- adopting an escalation process through which employees and others can raise concerns and seek guidance on the company's compliance with its policies.

Deployers might have similar, but different, governance obligations, including, for example, a requirement to take reasonable steps to avoid deployments of the AI system that could pose unacceptable risks of harm.

If the Commission does choose to go beyond governance requirements (as it does in the White Paper), it might be better to frame the proposed mandatory requirements as a non-exhaustive "toolkit" list of measures that regulated actors can use to promote trustworthy AI with respect to high-risk AI applications. That approach would provide developers and deployers with the needed flexibility to choose those safeguards that are most appropriate for the context.

Turning to the proposed requirements themselves, and specifically those relating to training data, record-keeping, robustness, and accuracy:

Training data

The Commission proposes, among other measures, that regulated actors "take reasonable measures so the use of AI systems does not lead to outcomes entailing prohibited discrimination" and "use data sets that are sufficiently representative, especially to ensure that all relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination are appropriately reflected in those data sets." These concepts are highly abstract and raise complex interpretive questions and implementation challenges. For example, how can developers determine whether their training data is "sufficiently representative"? Should they compare distributions in the training data to distributions in the population at large? Or only in those likely to be impacted by the AI (e.g., users on the platform)? Should they consider the population distribution only in the relevant sector (e.g., workforce in a given industry) or as a whole? What if a certain group is under-represented in the overall population (only 0.5% of the population is of a certain ethnicity)—should the training data mimic that under-representation?

Sometimes statistical techniques deliberately oversample some groups, especially if variances across groups are not equal, or if some groups had low participation. For example, imagine that a particular category contained only 1% women, with a sample size of 100 people—in such instances, a developer might ask for the training dataset to include more than one woman despite the fact that the dataset was "sufficiently representative" of the category.

These points are not intended to question the importance of having sufficiently representative training data—to the contrary, Microsoft researchers and product teams have collectively spent thousands of hours working on this issue. That work, however, has revealed the complexity of this issue: analyses are highly contextual; answers may vary depending on the goals of the AI system and the scenario in which it is used; and without common terminologies, frameworks, and tools, teams may reach different conclusions on what constitutes a sufficiently representative datasets. We share the goal of the Commission on this aspect, and we are committed to continuing to work to find a path forward.

Keeping of records and data

The Commission proposes, among other measures, that the regulation might require retention (and, possibly, provision to a regulator) of "accurate records regarding the data set used to train and test the AI systems, including a description of the main characteristics and how the data set was selected" and "in certain justified cases, the data sets themselves." It is unclear, however, what would constitute a "justified case" requiring retention / disclosure of a dataset or, in cases where a dataset included personal data, how those requirements would be reconciled with the GDPR (e.g., data minimization obligations). As set out in Annex II, Microsoft is working to develop tools that those involved in assembling or using training data can use to convey the key characteristics of the dataset while still protecting privacy and other important interests; our proposed "Datasheet for Datasets" is one result of these efforts. The value of such tools, however, depends in part on the broader governance framework; using them in isolation or after the fact is unlikely to be as effective in promoting trustworthy AI.

Robustness and accuracy

This requirement would include an obligation to ensure that "all systems are robust and accurate, or at least correctly reflect their level of accuracy, during all life cycle phases" and that "AI systems can adequately deal with errors or inconsistencies during all life cycle phases." The White Paper does not discuss how accuracy would be evaluated, or what the benchmark would be. Although those points are critical, it is equally important to recognize that there is no single, "correct" level of accuracy. For example, accuracy levels for an AI system used to decide when to apply the brakes on an autonomous vehicle would be meaningfully different than that used to predict whether a consumer would prefer a green or blue blouse.

Different accuracy levels may be appropriate even with regard to a single AI application—for instance, a facial recognition system used to allow consumers to access their bank accounts needs to be much more accurate than one used to "recognize" faces in a photo organizer app. Rather than mandate a specific level of accuracy or robustness, it might be more appropriate to require regulated actors to provide transparency about levels of accuracy during all phases of the AI system lifecycle. Additionally, the regulation could support the use of governance standards and procedures, as well as technologies, systems, and tools, as previously suggested.

Human oversight

The White Paper proposes various approaches to human oversight, which it indicates would vary depending on the "intended use of the systems and the effects that the use could have for affected citizens and legal entities." We agree that human oversight can be critical, and that it is often a necessary element to ensuring that an AI system is appropriately accountable, in particular where there are consequential impacts. How a given system is designed to operate, however, is an essential element in determining what type of human oversight is appropriate. Some AI systems operate at a scale where validating manually (through a human) every single AI output is simply not possible.

Operate at a scale where validating manually (through a human) every single AI output may not be possible.

Addressees of obligations

In discussing which actor in the AI lifecycle must comply with which requirement, the White Paper states that "each obligation should be addressed to the actor(s) who is (are) best placed to address any potential risks" (p. 22). We agree. As noted earlier, AI regulation should be clear on which requirements apply to which actors and ensure that responsibilities always fall on the actor that can most efficiently comply with them. This will be particularly important with regard to AI systems that have multiple possible applications, since developers in such scenarios may have only limited insight into the particulars of its deployment and limited control over how it is deployed (although ideally developers' principles for responsible AI would be built to apply throughout the AI system's lifecycle, including its deployment). Likewise, the compliance and enforcement regimes should align with this allocation of responsibilities and not require one set of actors to ensure compliance with requirements that reasonably should fall to another.

Compliance and enforcement

The White Paper proposes to enforce the future regulatory regime through mandatory pre-marketing conformity assessments.

As a company that develops and deploys a wide range of products and services, we have significant experience with these types of assessments in other sectors, such as cloud computing and computer hardware. And the EU, of course, maintains a range of conformity assessment approaches for products being placed on the market in the EU. It will be important for any future AI regulation to leverage, and learn from, these existing regimes, both with regard to best practices and things to avoid. Evaluations of current EU product conformity regimes (e.g., the Commission's 2017 evaluation of the Machinery Directive) might be useful to consider in this regard.

That said, it is worth emphasizing that the EU's current conformity assessment regime is built around products. The Commission should not underestimate the challenge of extending this regime to AI systems, which encompass a wide range of digital services. It will be hard work to ensure that conformity assessments of AI systems can be done quickly and efficiently. The regime should not lead to businesses and consumers in the rest of the world having access to AI innovations months or even years before those in Europe do.

Compliance of AI systems with the governance requirements proposed above should be self-assessed ex ante, managed responsibly throughout their lifecycle, and enforced ex post by authorities. Self-assessment is the most scalable approach. It is also the approach most appropriate for dynamic AI services, and the one that is least likely to unnecessarily extend time to market or unduly burden smaller operators. Equally important, self-assessment makes good sense in the context of AI systems, where the expertise to assess a system is most likely to lie in the organization itself.

In addition, international and European standards should be used as the basis of criteria used to assess compliance with any regulatory requirements.

This approach is commonly used by industry in Europe for cybersecurity (ISO/IEC 27001) and privacy (ISO/IEC 27701) assessment. Certification schemes based on industry standards is an efficient and common way to establish and maintain responsible business practices at scale across an entire economy. In addition, under the European Commission's "New Approach", adopting certain standards as "harmonized" enables certification of those standards to establish a presumption of conformity to regulation. Such harmonization also protects against inconsistent obligations between Member States that would result in unintended trade barriers or prohibitive costs for service offerings across the Union.

The ISO/IEC committee dedicated to AI (ISO/IEC JTC1 SC42) is currently developing international standards on AI foundational concepts (ISO/IEC 22989 and ISO/IEC 23053), on governance for the use of AI (ISO/IEC 38507), and on AI risk management (ISO/IEC 23894).

The ISO/IEC committee is starting the development of AI management systems standards which can become the criteria for the certification of responsible AI management practices. We also direct the Commission to the committee's recently published Technical Report (ISO/IEC TR 24028), which provides a helpful overview of topics relevant to building trustworthiness of AI systems. 19 European Member States are actively involved in this international standardization work.

In terms of ex post market surveillance, it will be important to consider carefully the best possible authority to oversee compliance. While national authorities may well need to play a role to enforce the rules locally and / or in specific sectors, pan-EU coordination will be critical given the cross-border nature of AI systems.

Voluntary compliance framework

While “high-risk” AI would be subject to mandatory obligations, the White Paper proposes a voluntary labelling scheme for all other AI systems. Microsoft sees value in such an approach. Many companies, including Microsoft, are working hard to adopt policies and procedures to promote trustworthy AI. Labelling could prove beneficial for end-users when comparing and choosing AI solutions. Signaling to the market that a given AI systems meets a defined set of requirements for trustworthy AI will also help leverage the power of supply and demand to incentivize companies to adopt responsible practices.

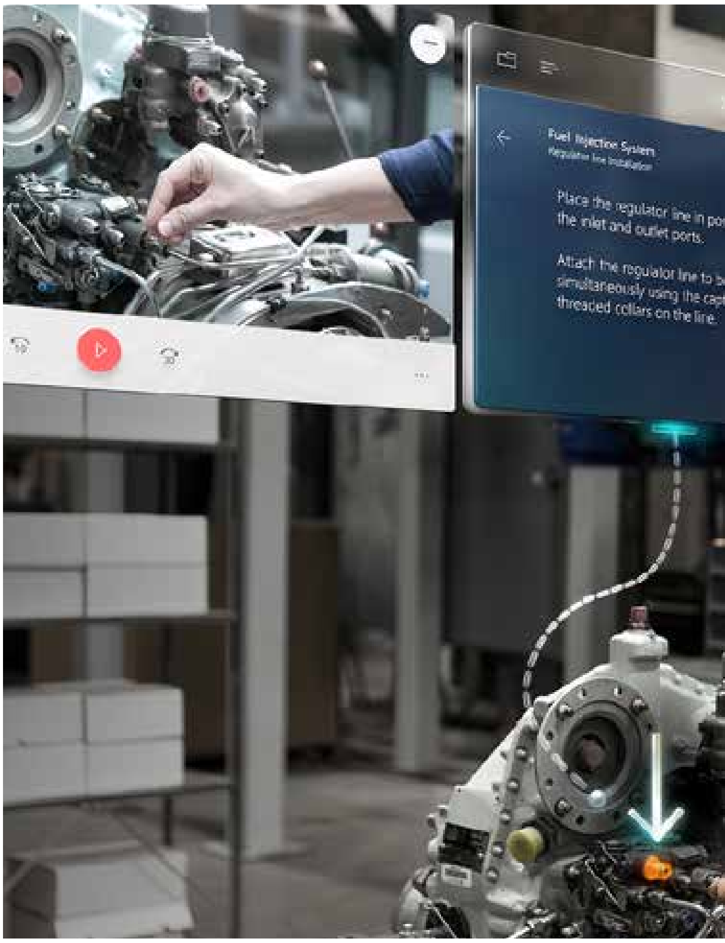
Any voluntary labelling scheme should be designed for all types of organizations, including SMEs, to encourage the greatest participation. Labels should be used to demonstrate adherence to good governance processes (e.g., “this AI system was developed by a manufacturer that meets [formal standard] [named governance best practice]”) rather than outcomes (e.g., “this AI system is accurate”). In addition, given the diverse range of AI solutions on the market, and the need to keep any labelling regime efficient and cost-effective, labelling should not be one-size-fits-all. Instead, there should be a common set of high-level criteria to ensure that the labels are well recognizable and understood by customers and end-users. Developers and deployers should have the ability to self-assess their products and services against the requirements of any future regulation, and against relevant and recognized European or international standards or industry codes. When they meet the relevant requirements, they should be able to assert this on their products / services in a manner that is consistent with the common labelling criteria.

Asserting compliance with the requirements of any future regulation should not subject a developer or deployer to the same ex ante and ex post enforcement regime that applies to high-risk AI, however. Extending the regime in this way would deter many organizations from choosing voluntarily to comply. Instead, organizations that falsely assert that their products or services are “compliant” should be subject to penalties.

Safety and Liability

The White Paper also suggests that reforms of the EU’s product safety and liability regimes may be necessary to address AI and other emerging technologies. Rules on product safety and product liability serve complementary purposes—one seeks to protect consumers from harm ex ante, and the other applies ex post, ensuring compensation where harm occurs.

Given this relationship, considering the regimes holistically makes sense. But we recommend waiting to undertake that review until the parameters of the AI safety regime are settled, so that the liability rules reflect the safety framework (see also comments on robustness and accuracy).


Fuel Injection System
Regulator line Installation
Place the regulator line in po
the inlet and outlet ports.
Attach the regulator line to b
simultaneously using the cap
threaded collars on the line.

As an initial step, we support the Commission's evaluation of product safety legislation in the context of its consideration of broader AI regulation. Given that the regime proposed in the White Paper would encompass "safety" in the broadest sense (i.e. in terms of material and immaterial harms and harms to fundamental rights), it will be important to ensure this regime aligns with existing EU safety frameworks, legislation, and standards. Safety requirements should complement cybersecurity requirements in particular, to help ensure that AI components remain robust and resilient despite an attack.

A product safety evaluation can include an assessment of cybersecurity risks and whether the product conforms to cybersecurity requirements that would mitigate other safety issues such as system robustness. Any consideration of AI product safety and security requirements should be done in coordination with relevant EU cyber entities, such as the European Network and Information Security Agency ("ENISA") and the proposed European Cybersecurity Network and Competence Centre. Requiring duplicative assessments will be unnecessarily burdensome (in particular for SMEs), and also less efficient from a risk management point of view.

In terms of reform to the EU's liability framework, Microsoft endorses the starting point that users deserve the same liability protections for products that incorporate AI as for all other products. The mere fact that a product includes AI technology, however, should not mean that it is subject to stricter liability rules. The current EU product liability regime, established in the Product Liability Directive ("PLD") and complemented by national liability regimes, has worked well in a wide variety of contexts for over 30 years.

During much of this time, PCs, laptops, smartphones, and other digital devices have been ubiquitous, yet there is little evidence of consumers being unable to recover for harms caused by defects in these products. Before considering major changes to the PLD, further empirical evidence is needed to identify those cases where parties have been unable to obtain redress under the existing regime and to understand the reasons for that inability.

The same approach should be taken to any EU-level changes that will affect Member State regimes. These regimes have been carefully crafted to fairly apportion evidentiary burdens between plaintiffs and defendants. Changing them—e.g., by imposing a presumption of fault against AI-driven products—could undermine incentives for investment in the AI innovation, and should not be considered absent clear empirical evidence that there is a need to do so.

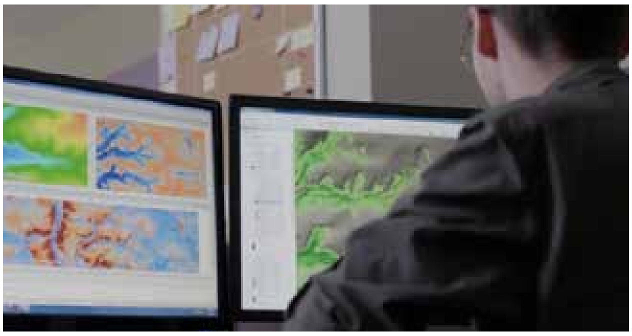
AN ECOSYSTEM OF EXCELLENCE

As we continue to develop innovative new AI technologies, Microsoft is committed to working in partnership with Europe's public sector, universities, civil society, and industry to promote AI applications that align with European fundamental rights and values. We agree that creating an "ecosystem of excellence" around AI will require coordinated investment in research and innovation, skills and talent, and market uptake among small and medium-sized enterprises ("SMEs") and the public sector. We support efforts to empower European institutions, large and small, to drive economic and social progress and compete on a global scale.

Microsoft's operations throughout the EU already are working to achieve many of the goals set out in the White Paper. Below we offer a snapshot of just a few or our many efforts and initiatives that support these critical objectives.

Focusing the efforts of the research and innovation community

The EU is uniquely suited to the task of leading the world in AI research and innovation. Europe's diversity of cultures, traditions, and perspectives all serve as tremendous assets in building and maintaining a cutting-edge research and innovation community. This powerful advantage will allow Europe to develop and leverage unique AI-driven solutions to address major societal challenges, including climate change, access to health care, and the modernization of public services. Microsoft currently supports EU research and innovation through multiple initiatives and is committed to continued investment in this area.



Microsoft Research

Microsoft Research has established itself as a premier research organization, contributing to the advancement of computer science and the development of Microsoft products and services. Microsoft Research broadly shares many of its findings, allowing AI researchers throughout the EU and the world to build on these advances and insights. Our researchers have published more than 22,000 papers in all areas of study, including critical fields such as environmental science, health, privacy, and security. Microsoft Research is truly global in scope, including through active research centers in Europe.

Zurich Lab

The Microsoft Mixed Reality & AI Zurich Lab is focused on building the future of mixed reality. In collaboration with the Swiss Federal Institute of Technology and other top technical schools in Europe, the Zurich Lab is training the next generation of AI researchers in the EU, providing them with the tools and experience necessary for a career in AI.

Cambridge Lab

The Microsoft Research Lab in Cambridge adopts an interdisciplinary approach to AI research, developing models and techniques that can be applied to complex real-world data sets.

In partnership with the University of Cambridge, the Cambridge Lab trains and supports PhD students, postdoctoral researchers, and interns, allowing them to develop tools that enhance the human experience.

Microsoft-Inria Joint Center

Since its founding in 2006, the Microsoft Research – Inria Joint Center has brought researchers together to apply computer science and mathematics principles to a range of scientific challenges, including machine learning and big data. They have recently focused their joint efforts on accelerating the deployment and adoption of AI within France's technology ecosystem.

Microsoft Research AI

The Microsoft Research AI (MSR AI) initiative brings together the breadth of talent across Microsoft Research to pursue critical advances in AI. This research and development initiative combines advances in machine learning with innovations in language and dialog, human-computer interaction, and computer vision to address some of the toughest challenges in AI.

Innovation

> **"The future is discovered in the unexpected. It's found where no one has thought to look. And where no one, yet, has had the courage to go."**
>
> Microsoft

For Microsoft, true innovation hinges on the ability to see things differently, which is why we believe that Europe's diversity of perspectives, backgrounds, and experiences will be prove to be an enormous strength in building the EU's research and innovation community. Microsoft strives to tackle large, global issues through AI innovation with its European partners, and would be eager to formalize and expand its collaboration with public and private entities across the EU.

Examples of Microsoft initiatives that are working to use AI to solve major challenges include the following:

AI for Good. This **initiative** provides technology, resources, and expertise in support of organizations and individuals working to address critical issues facing society in five areas: environment and sustainability, healthcare, accessibility, humanitarian action, and cultural heritage.

AI for Health leverages AI tools to advance medical discoveries, uncover global health insights, and to increase health equity across underserved populations. For example, Microsoft has partnered with the Novartis Foundation to develop an AI-enabled tool for early detection of leprosy. We have also mobilized AI for Health in response to the coronavirus pandemic, addressing issues ranging from treatment and diagnostics to allocation of resources.

AI for Earth provides AI technology and cloud software to those working to solve global climate issues. For example, the German company Breeze Technologies received an AI for Earth grant to support its efforts to improve air quality with AI technology, and an engineer and researcher at the National University of Ireland Galway is using Microsoft AI technology provided through AI for Earth to study the decline of bee populations.

A Planetary Computer

This story shares how through AI the power of cloud computing we can convert massive amounts of data into the insights and information needed by researchers, academics and environmental experts to create a more sustainable planet, redefining areas such as agriculture, biodiversity, climate change and water. To protect nature and its biodiversity we must monitor and analyze its health. To draw the full picture, we need data that currently data scientists are lacking of as collecting data is a manual and labour-intensive process. To overcome this obstacle, Microsoft is building a Planetary Computer powered by Azure Open Data to help scientists from around the world to monitor, model and manage the planet's natural resources. It is our aim to make available a massive treasure trove of data that will be accessed and updated by anytime, anyone, anywhere on the planet. It is our vision that this data will become actionable insights and will tell us how to prioritize, preserve and protect animal and marine species, forests, waterways and ecosystems around the world.

AI for Accessibility supports independence and productivity among the disabled population, creating AI solutions in three key areas: employment, daily life, and communication. For example, Microsoft partnered with the French company Equadex to create a pictogram app for nonverbal children using technologies developed by Microsoft Cognitive Services and Microsoft's Azure cloud platform.

AI for Humanitarian Action uses AI solutions to support disaster recovery, address the needs of children, protect displaced populations, and promote human rights worldwide. For example, Microsoft recently collaborated with the Norwegian Refugee Counsel, NetHope, and the University College of Dublin to create a chatbot that uses AI technology to connect youths with online learning resources when they do not have access to schools.

AI for Cultural Heritage aims to use AI for the preservation and enrichment of cultural heritage. For example, AI for Cultural Heritage brought together the Musée des Plans-Reliefs, Microsoft, and French technology companies to animate the Mont-Saint-Michel relief map through AI and mixed reality.

iPrognosis

This project, which received funding from the European Union's Horizon 2020 research and innovation program, applies technology solutions to diagnosing and managing Parkinson's disease. This consortium includes eleven organizations from 6 EU countries, including the Microsoft Innovation Center in Greece.

Automatic Speech Recognition

Microsoft's Azure Speech Service provides speech recognition, voice fonts, and speech translation in over 30 languages, including 12 EU languages. In addition, Microsoft Translator translates between 60 languages, including all 24 EU languages. The European Parliament uses these technologies to translate parliamentary debates in real time, and the Commission is using Microsoft's open-source technology for its translation needs.

Skills

In order to create an ecosystem of excellence for AI in Europe, it is absolutely critical to ensure that the EU workforce has the necessary technical skills. This requires creating an education and training pipeline for AI experts, in addition to improving basic knowledge and technical skills of European workers generally. To that end, Microsoft is committed to supporting employers, jobseekers, and students in developing the skills necessary to succeed in the AI age. We have partnered with the public sector, academia, industry, and civil society in numerous efforts aimed at developing AI educational and skill-building opportunities throughout Europe, including the following:

Ambizione Italia #DigitalRestart

This initiative is a five-year commitment to invest 1.5 billion USD in support of innovation and growth in Italy through training, digital transformation, and open innovation efforts. As part of this initiative, Microsoft will create AI hubs to help Italian citizens and businesses develop and deploy AI solutions in retail, design and fashion, manufacturing, financial services, healthcare, and infrastructure through dedicated grants and resources.

AI School in Belgium

In an effort to address the shortage of AI experts in Belgium, Microsoft worked with several partners to launch a first-of-its-kind School of Artificial Intelligence in the country. The goal is to open nine schools across Belgium, which will train up to 500 people every year. To ensure diversity and accessibility, students may attend the school tuition-free.

Microsoft Azure Academy in the Netherlands

The Microsoft Azure Academy aims to expand the pool of female engineering and IT talent in Europe, particularly in cutting-edge fields such as cloud computing and big data. Those who complete the program are qualified to work as junior data scientists.

Initiatives for Digital Education and Skills in Germany

Microsoft launched the award-winning Initiatives for Digital Education and Skills in Germany, offering a range of technical skills and training programs. This includes Code Your Life, which introduces children to computer programming and AI, and IT Fitness, which provides self-directed training on AI and automation to youth and adults alike.

Focus on SMEs

SMEs play a critical role in the European economy, employing 87 million people and contributing to the region's competitiveness. Microsoft has a longstanding history of supporting SMEs and is committed to continue doing so, in partnership with the Commission and other European stakeholders. Examples of Microsoft efforts to support SME success include the following:

Microsoft ScaleUp

The [Microsoft ScaleUp program]() is designed for Series A-C startups, offering access to sales, marketing, and technical support for their growing businesses. Eligible startups take part in an immersive program at one of our eight global

Microsoft Reactors

[Microsoft Reactors]() are community spaces where developers and startups can meet, learn, and connect with local peers while also gaining access to industry-leading ideas and technology from Microsoft, partners, and the open source community. Microsoft Reactor spaces can be found across the world, including in Stockholm and London.

Global Social Entrepreneurship Program: This program, launched in 2020, supports social impact startups as they build and scale globally. It is available in 140 countries, and actively seeks to support underrepresented founders with diverse perspectives and backgrounds. The criteria to qualify for the program include a business metric that measures impact on an important social or environmental challenge; an established product or service that will benefit from access to enterprise customers; and a commitment to the ethical and responsible use of AI.

Partnership with the private sector

Public-private partnerships are critical to building an AI ecosystem of excellence in Europe; the private sector brings deep technical knowledge and innovation, and the public sector provides unique resources and direction. In addition to the public-private partnerships described above, Microsoft has participated in a number of consortia and other initiatives with governments throughout Europe. A few recent examples of these efforts include the following:

AI4Belgium Coalition

AI4Belgium is a community-led approach to enable Belgian people and organizations to capture the opportunities of AI in a responsible way. The coalition includes representatives from academia, public institutions, and industry, including start-ups. The Coalition's goals is to assist Belgian policymakers in ensuring that AI becomes a positive force for Belgium and that the country has the necessary skills to realize AI's full potential.

Impact AI

Impact AI is the first French think tank dedicated to AI issues, bringing together technology companies, private and public entities, research institutes, and academics. Launched in March 2018, Impact AI seeks to accelerate positive impact of AI by reflecting on ethical and societal challenges of AI. It also supports innovative projects that will help demystify AI among the French people and promote trustworthy and ethical AI in France.

Promoting the adoption of AI by the public sector

Microsoft agrees that the public sector could benefit enormously from the adoption of AI solutions. In our decades-long work with governments around the world, we have seen that an open mind about new technologies and willingness to change is critical to the modernization of public institutions, and we have seen this hold true in the context of AI as well. A small sampling of our many efforts to work with governments to promote the beneficial use of AI include the following:

Combating Child Sexual Exploitation with AI

In partnership with the Ministry of Justice of North Rhine-Westphalia, the Cologne Public Prosecutor's Office, and Saarland University, Microsoft Germany is developing AI to identify and preserve evidence of child sexual exploitation in criminal cases. This technology aids law enforcement by allowing investigators to quickly evaluate seized evidence and determine whether it can be used in court.

Swedish Forest Agency

. With the support of an AI for Earth grant, the Swedish Forest Agency and Microsoft have worked together to develop and deploy AI solutions to anticipate and prevent damage to Sweden's forests. The Agency plans to make its pilot data, and the finished AI solution, available to third parties to enable them to identify and counteract infestations in new areas around Europe and the world.

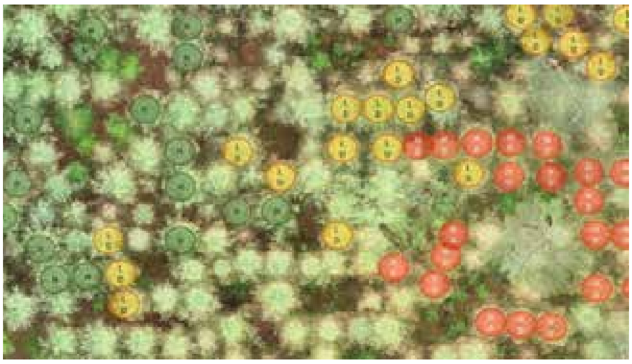


Illustration: Skogsstyrelsen. With the help of drone images, individual trees are marked to learn the model, identify all larch trees, and classify them as either healthy or damaged

European Global Fishing

Microsoft is working with the Commission's Directorate-General for Maritime Affairs and Fisheries to gain insight into fishing activity data through the use of AI. The goal of this effort is to help officials quickly identify suspect and potentially illegal fishing activities, and thereby to help preserve biodiversity in the oceans.

Securing access to data and computing infrastructures

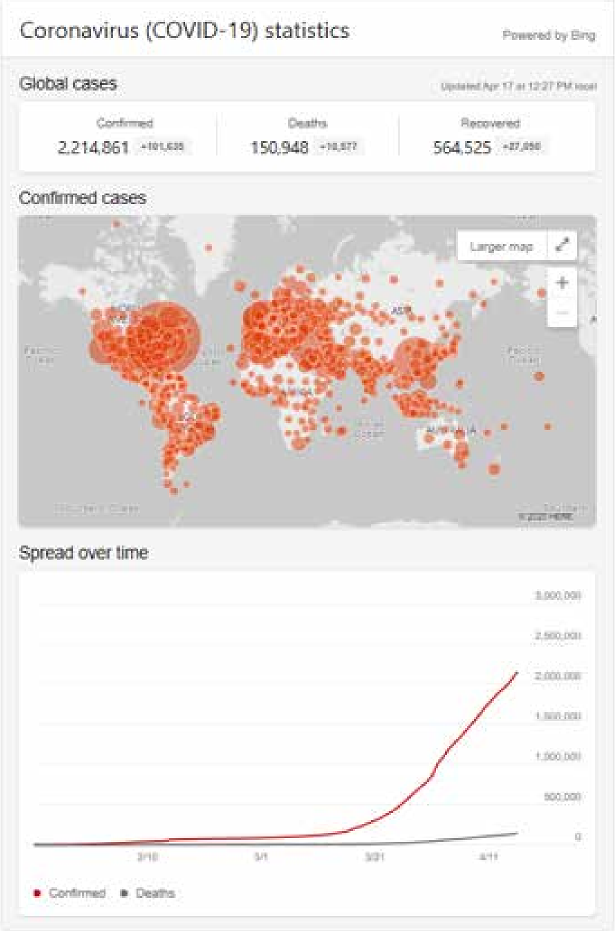
Recent events, including in particular efforts to respond to the COVID-19 pandemic, have demonstrated once again the benefits of promoting access to data. We therefore strongly support the Commission's desire to facilitate access to data across organizational boundaries, while also ensuring that data remains safe and secure and is fully consistent with EU laws and values, including its rules on privacy and data protection.

To promote the many benefits of data access, Microsoft recently launched an Open Data Campaign to help organizations of all sizes realize the benefits of data and the new technologies that data powers. We believe everyone can benefit from opening, sharing, and collaborating around data to make better decisions, improve efficiency, and help tackle some of the world's most pressing societal challenges. To help guide our own efforts on open data, we adopted a set of principles to inform how Microsoft can most effectively open and share data in a responsible way.

In addition to charting a principled course, we believe success will depend on building deep collaborations with others from across industry, government, and civil society around the world. To this end, Microsoft is committing to launching 20 data collaborations by 2022. Among other topics, our initial work will focus on the following:

Addressing COVID-19.

Recognizing that COVID-19 represents one of the most pressing challenges facing the world today, we are actively contributing to the work being done to use data to respond more effectively to the crisis. This includes our partnership with Adaptive Biotechnologies to decode how the immune system responds to COVID-19, which will share research findings via an open data access portal for any researcher to use in the fight against the pandemic. Microsoft also provides a COVID-19 tracker on our Bing search engine and is releasing aggregated data to those in academia and research. In addition, our GitHub team is hosting a range of collaborative COVID-19 projects, including open source software, hardware designs, models, and many leading COVID-19 datasets.



Coronavirus (COVID-19) statistics

Powered by Bing

Global cases

Updated Apr 17 at 12:27 PM local

Confirmed	Deaths	Recovered
2,214,861 +101,635	150,948 +10,577	564,525 +27,050

Confirmed cases

Spread over time

Helping cities collaborate around data

Microsoft will partner with Arup and the Oliver Wyman Forum on the London Data Commission, an open data initiative run by London First working with the Greater London Authority and others, to lead a data collaboration project around city-based data that can help address social and economic challenges in London.

Advancing data-driven healthcare

Microsoft is working with the Novartis Foundation, Apollo Hospitals in India and Coala Life in Sweden to consolidate their respective cardiovascular datasets from hospitals and primary care centers around the world. This will be one of the first global data collaboratives to improve cardiovascular health, bringing together data from a range of sources to help address one of the world's leading causes of death. The collaborative aims to further develop and use the leading cardiovascular AI tool–AICVD Risk Score, created by Apollo Hospitals–to accelerate the use of data-driven decisions in tackling cardiovascular disease.

International aspects

The EU's regulation of, and investments in, AI has the potential to become a global model for how governments support and promote trustworthy AI. When engaging with the international community, Microsoft urges the Commission to focus on promoting regulatory convergence and supporting multi-stakeholder efforts to advance, implement, and make available AI solutions.

Microsoft actively encourages international cooperation on the promotion of AI through multiple global initiatives—examples of these efforts include the following:

Industry and Multi-Stakeholder Initiatives

Microsoft has participated in several international multi-stakeholder initiatives on AI. This includes the Partnership on AI, the European AI Alliance, the OECD's Expert Group on AI in Society, the 2016 White House consultation on Preparing for the Future of Artificial Intelligence, the Singaporean Government's Advisory Council on the Ethical Use of AI and Data, the AI4People Forum of the Atomium European Institute for Science, Media, and Democracy, the AINow Institute at New York University, the ISO/IEC JTC 1/ SC 42 AI standardization efforts, and many others.

OECD AI Policy Observatory

After the OECD released its AI Principles—the first intergovernmental standard on AI—it launched the OECD Policy Observatory to help governments develop trustworthy AI to benefit society. Microsoft and LinkedIn are contributing to this effort through the Microsoft Academic Graph (with valuable data on scientific publications and research networks) and the LinkedIn Economic Graph (with skills-development and migration insights), which together are designed to help advance more informed national AI policies and regulations. We are also participating in the OECD Network of Experts on AI (ONE AI) and are co-leading a working group on the implementation of trustworthy AI.

Rome Call for AI Ethics

Together with the Pontifical Academy for Life, IBM, FAO, the Italian Government, and the Vatican, Microsoft signed the Call for AI Ethics, a document developed to support an ethical approach to AI and promote a sense of responsibility among organizations, governments, and institutions around the globe as we embrace AI solutions in our everyday lives.

Microsoft's Journey To Responsible AI

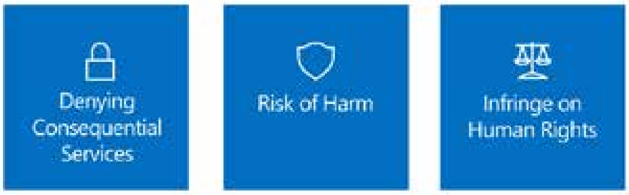
Although we are continually amazed at the innovative ways in which our customers and partners are discovering new ways to use AI to benefit people and society, we also fully appreciate that AI systems, if not designed carefully and deployed appropriately, can create risks. Our recognition of these risks led Microsoft, in 2016, to announce a set of six principles that would guide our development and deployment of trustworthy and ethical AI, which we further developed in The Future Computed, released in 2018. Since that date, we have been on a journey to refine those principles, operationalize them in our work, and develop new technologies, systems, and tools to help both our internal teams, and our partners and customers, develop and deploy AI systems in a more responsible and trustworthy manner. It has been a journey of experimentation, learning and culture change, and one on which we still have far to travel.

> "Technology companies must take responsibility for what they create and the impacts that their technology have on the world."
>
> Brad Smith, President, Microsoft

The six principles that govern Microsoft's development and use of AI are: fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability. These principles—which align to a large extent with those set out by the Commission's High-Level Expert Group on AI ("AI HLEG") in its Ethics Guidelines for Trustworthy AI—guide our "full lifecycle" approach to responsible AI, from its design and development through to its deployment.

Of course, principles have value only if they are lived by. In our experience, robust governance standards and processes are vital to operationalizing responsible AI. Our governance mechanisms include robust internal policies that help us to ensure our development and deployment of AI reflect our principles. We have also established two bodies to help us put our AI principles into practice: Microsoft's AI, Ethics, and Effects in Engineering and Research (Aether) Committee, and our Office of Responsible AI. Aether is tasked with advising Microsoft's leadership around questions, challenges, and opportunities brought forth in the development and fielding of AI innovations. The Office of Responsible AI endeavors to implement our cross-company governance, enablement, and public policy work. Together, the Aether Committee and the Office of Responsible AI work closely with our engineering and sales teams to help them uphold Microsoft's AI principles in their day-to-day work.

One key learning we have gained is that making progress on responsible AI requires more than just principles and governance procedures—it also requires 'tools' that people who are designing and deploying AI can use to identify and mitigate risks.



To that end, Microsoft is heavily engaged in developing and testing standards, frameworks, and processes for the development and use of responsible AI systems. This work builds on Microsoft's long history of innovation to make computing more accessible and dependable for people around the world—including the creation of the Microsoft Security Development Lifecycle and pioneering work in accessibility and localization. We provide more information about our work in this area in Annex II.

In addition to the work we are doing internally to promote responsible AI, we are also working closely with customers, other tech companies, academia, civil society, governments, and many others. As one example, Microsoft is a co-founder of the Partnership on AI ("PAI"), a working group that brings together industry leaders, academics, non-profits, and specialists to develop best practices, and provides an open platform for discussion and engagement around AI's impact on people and society. Recent initiatives by PAI include a research project to explore limitations on access to and usage of demographic data as a barrier to detecting bias in AI systems. Additional examples of Microsoft's participation in such multi-stakeholder efforts are described in our response to the Commission's proposals on an Ecosystem of Excellence above.

Resources For Putting Responsible AI Principles Into Practice

Researchers at Microsoft are making significant contributions to the advancement of responsible AI practices, techniques, and technologies. These contributions span the areas of human-AI interaction and collaboration, fairness, intelligibility and transparency, privacy, reliability and safety, and other areas. We have made many of our tools—including relevant research papers, open source projects, and Azure Machine Learning tools—available to the public in order to aid other developers and data scientists in understanding, protecting, and controlling their AI systems.

Fairness

For those seeking to prioritize fairness in their AI systems, Microsoft researchers, in collaboration with Microsoft's Azure Machine Learning team, developed Fairlearn, an open-source Python package that helps developers of AI systems assess whether the system might have negative impacts on groups based on factors such as race, gender, age, or disability status. Fairlearn, which focuses specifically on harms of allocation or quality of service, draws on two papers by Microsoft researchers on incorporating quantitative fairness metrics into classification settings and regression settings, respectively. Of course, even with precise, targeted software tools like Fairlearn, it is still easy for teams to overlook fairness considerations, especially when they are up against tight deadlines. This is especially true because fairness in AI sits at the intersection of technology and society, and cannot be addressed with purely technical approaches. To assist in this effort, Microsoft researchers have therefore co-designed a fairness checklist that is intended help teams reflect on their decisions at every stage of the AI lifecycle.

Intelligibility

For those eager to incorporate intelligibility into their own AI system lifecycles, Microsoft researchers have released InterpretML, an open-source Python package that exposes common model intelligibility techniques to practitioners and researchers, allowing them to better understand and debug their machine learning models. InterpretML includes implementations of both "glassbox" models (like Explainable Boosting Machines, which build on Generalized Additive Models) and techniques for generating explanations of blackbox models (like the popular LIME and SHAP, both developed by current Microsoft researchers). Beyond model intelligibility, a thorough understanding of the characteristics and origins of the data used to train a machine learning model can be fundamental to building more responsible AI.

Data transparency

The Datasheets for Datasets project proposes that every dataset be accompanied by a datasheet that documents relevant information about its creation, key characteristics, and limitations (e.g., why the dataset was created, what information it contains, and the tasks for which it should and should not be used). Datasheets can help dataset creators uncover possible sources of bias in their data or unintentional assumptions they have made, help dataset consumers figure out whether a dataset is right for their needs, and help end users gain trust. In collaboration with the Partnership on AI, Microsoft researchers are exploring how to develop best practices for documenting all components of machine learning systems to build more responsible AI.

Accuracy and robustness

To help developers and deployers understand the way reliability and safety problems may occur in AI systems, Microsoft researchers have been investigating how blind spots in data sets, mismatches between training environments and execution environments, distributional shifts and problems in model specifications can lead to shortcomings in AI systems.

Given the various sources for failures, the key to ensuring system reliability is rigorous evaluation during system development and deployment so that unexpected performance failures can be minimized and system developers can be guided for continuous improvement. That is why Microsoft researchers have been exploring avenues for developing new techniques for model debugging and error analysis that can reveal patterns that are correlated with disproportionate error regions in evaluation data.

We recognize that when AI systems are used in applications that are critical for our society, in most cases to support human work, aggregate accuracy is not sufficient to quantify machine performance. Researchers have shown that model updates can lead to issues with **backward compatibility** (i.e., new errors occurring as a result of an update), even when overall model accuracy improves, which highlights that model performance should be seen as a multi-faceted concept with human-centered considerations.

We embrace open collaboration across disciplines to strengthen and accelerate responsible AI, spanning software engineering and development to social sciences, user research, law and policy. To further this collaboration, we open-source many tools and datasets that others can use to contribute and build upon.