IBM Submission to the European Commission on the Draft Artificial Intelligence Act

IBM firmly believes in developing and helping others build responsible AI. Back in 2018 we put forward Trust and Transparency principles[1] which include key commitments to develop and apply explainable and transparent AI with a purpose to augment – not replace – human intelligence.

We have put these principles into practice[2] through different initiatives like the development of trustworthy AI open-source toolkits[3] to support our clients' journey to trustworthy AI, and establishing a cross-disciplinary AI Ethics Board. Through our contributions to the European Commission's High-Level Group on AI Ethics, the signature of the Rome Call for AI Ethics, our strong commitments on the ethical use of technology[4] and our many contributions to the discussions around AI policy[5], we have endeavored to show that technology should be, and can be, developed and deployed in a responsible way without hindering innovation.

Building trust requires acknowledging valid concerns that exist in relation to accountability, transparency, fairness and security, and putting in place appropriate regulatory mechanisms to manage those risks, while continuing to promote ongoing innovation and experimentation – getting that balance right requires a precision regulation approach that is clear and targeted.

General considerations on the draft Artificial Intelligence Act

We commend the Commission for the continuity and coherence of the 2019 White Paper on AI and the draft Regulation, and for following key recommendations of the Commission's High Level Expert Group on AI Ethics in drafting the legal requirements for high-risk AI systems.

We believe the draft rules are built on the right set of principles and we support a risk-based approach to foster trust in AI without hindering its responsible development. We have long called for "precision regulation" through a proportionate approach that would regulate high-risk use cases, not the AI technology itself. We have supported the idea that AI systems should be considered high-risk when they meet certain criteria such as the severity and the probability of occurrence of certain serious harms to persons.

While we welcome the risk-based approach, we believe that some adjustments and clarifications will be useful to make sure the new rules are proportionate, legally clear and risk-based, so that implementation of the regulation and its practical effect on AI systems is consistent.

[1] https://www.ibm.com/blogs/policy/trust-principles/
[2] https://www.ibm.com/artificial-intelligence/ethics
[3] https://aif360.mybluemix.net/
[4] https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/
[5] https://www.ibm.com/blogs/policy/ai-precision-regulation/

Definition of AI systems

We support the Commission’s objective to make this regulation future-proof and technology neutral and understand that the rules are triggered only if a system, through its intended use, classifies as involving a risk of impact to health and safety or fundamental rights.

This being said, we have two fundamental comments about the definition of AI systems proposed by the Commission.

AI *tools* (as distinct from AI *systems*) often have no broader purpose beyond serving as building blocks for various user-designed applications, which in turn serve more specific user-generated intended purpose. An AI tool for which the "human-defined objective" is best characterized as "assisting in the development of user-designed AI systems" should not fall within the definition of AI systems, even if they use the techniques listed in Annex I and are capable of generating outputs that, broadly speaking, influence the environment. For instance, IBM Watson Studio[6] is an AI tool consisting of a workspace that includes multiple collaboration and open-source tools for use in data science that can be used as an AI building block. Some examples of such building block tools included in IBM Watson Studio provide the capability to operationalize AI models and manage their robustness and fairness in an automated manner. Consistent with Recital 60, we do not think it is the Commission’s intention to cover software, tools and models which are not in and of themselves AI systems.

Furthermore, the definition of AI systems in Article 3.1 and the list of techniques and approaches in Annex I give the Regulation a very broad scope and could potentially cover a significant amount of software applications which are not traditionally considered as AI technology. For instance, many applications which do not use AI technology generate outputs influencing the environments they interact with, such as a GPS system.

The draft Regulation aims for a framework whereby entities that procure an AI tool and turn it, through development and training, into AI systems intended for high-risk use, are responsible for compliance with the requirements for high-risk systems (Article 28). In this setting, the providers of general-purpose AI tools will cooperate with the purchasers of these AI tools and thereby support them in achieving compliance (Recital 60). Nevertheless, we believe that the definition of AI systems should be made more explicit regarding this distinction between general-purpose AI tools and AI systems.

For the reasons outlined above, we suggest small changes to the definition of AI systems to clarify that (i) general-purpose "building block" or "development" AI tools which serve as components or precursors of AI systems are not covered; and (ii) only software that genuinely uses AI technology fits in the definition.

In addition to expressly carving out general purpose tools, the definition of AI systems in Article 3.1 could be further adapted as follows: "*‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a ~~given set of~~ human-defined ~~objectives~~ intended purpose, generate outputs (such as content, predictions, recommendations, or decisions) influencing the environments they interact with and which cannot be fully predicted by the natural person developing the system.*"

[6] https://www.ibm.com/cloud/watson-studio

Unacceptable risk and prohibited systems

As a responsible technology innovator in developing AI-infused technologies, we agree with the Commission that the Regulation should ban certain practices which can lead to misuse of AI, and we therefore support the prohibition of the practices outlined in Article 5. We firmly oppose the use of technologies including facial recognition for mass surveillance, racial profiling and violations of basic human rights and freedoms – IBM was the first global company to publicly announce that we would no longer offer general purpose facial recognition or analysis software[7].

Low risk AI systems

As announced in our 2018 IBM Principles for Trust and Transparency[8], we believe that trust and transparency go hand in hand when it comes to the use of new technologies, including AI. We support the minimum transparency obligations of Article 52 for AI systems intended to interact with natural persons: people have the right to know when they are interacting with an AI system.

AI systems used in employment relationships

We believe that the use of AI systems in the area of employment overall does not per se qualify an application as high-risk. Our concern when reviewing Annex III is that the classification of all HR applications as high-risk does not recognize the need to differentiate between applications in the area of HR according to actual risks they pose to fundamental rights, and that such a wholesale categorization will stifle innovation in the near future and pose as a barrier to transforming HR processes to the benefit of workers and employers alike.

Consequently, IBM recommends delineating HR applications based on use case so that not every HR application is categorized as high risk.

Requirements for high-risk AI systems

Some requirements for high-risk AI systems, such as the ones referred to in Article 10.3 seem rather generic ("data sets must be relevant, representative […] and complete") or impossible to comply with ("data must be free of errors"). Also, the requirement to put in place human oversight that enables the user to "fully understand the capacities and limitations of the AI system" in Article 14.4.(a) is not possible to achieve in practice, since a developer cannot guarantee what a user will understand.

We understand that the articles describing the requirements for high-risk systems must be read considering Article 8 (Compliance with the requirements) and Article 9 (Risk management system) in particular, which states that "risk management measures (…) shall take into account the generally acknowledged state of the art, including as reflected in relevant harmonised standards or common specifications". If the Commission's idea is indeed that compliance with the requirements must be done with state-of-the-art levels in mind, as in "consistent with industry standards", then we believe the text could be amended to clarify this point further. In the case of data sets for example, it would be clearer to state that the "training, validation and testing data sets shall be relevant, representative, and appropriately vetted for errors and completeness in accordance with industry standards".

[7] https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/
[8] https://www.ibm.com/blogs/policy/trust-principles/

Obligations of providers and users of high-risk AI systems - General purpose AI tools

We urge the Commission to take note of the fluidity and diversity of roles and responsibilities in the AI marketplace, which is likely to continue as the relevant stakeholders experiment with technical development and business models to maximize the benefits of AI for society. For example, the purpose of an AI system may be dictated by a consumer-facing corporate user or the developer from which it purchases the system. A system may be developed entirely by one party, or it might encompass AI tools purchased from multiple vendors and patched together by a systems integrator or the user's employees. Training data may be obtained by a user internally or from vendors, and a system's sensitivity to new training data over time, or lack thereof, might dictate that the original developer or a later user should test for bias and accuracy. The methodology for identification of the party holding "provider" or "user" responsibilities should be clarified with a complex, evolving AI ecosystem in mind, in order to avoid stifling innovation for the benefit of all by providinglegal uncertainty.

Of particular concern for IBM, many of our products are general-purpose tools and APIs which customers use to develop and train, with their own data, AI systems across various industry sectors.

The intent of the draft Regulation is that a party using a general-purpose tool to develop and train an AI system for a high-risk intended use, is responsible for compliance with the requirements for high-risk systems (Article 28) – and providers of the general-purpose tool need to cooperate with users to support them in their compliance efforts (Recital 60). Nevertheless, we believe that the text of the Regulation must be more explicit regarding this allocation of responsibilities when it comes to general purpose tools.

In view of the above, we recommend to amend Article 28 on the obligations of third parties to (i) include in Article 28(1)(a) that users or other third-parties shall be considered a provider for the purposes of the Regulation when they place on the market or put into service a high-risk AI system including by training or otherwise modifying an existing AI system such that it becomes a high-risk AI system; and (ii) building on Recital 60, add as a new Article 28.3 that "third parties and notably the ones involved in the sale and the supply of software, software tools and components, pre-trained models and data, or providers of network services should not be considered providers for the purposes of this Regulation".

Standards, Conformity Assessment, Certificates and Registration

While we understand that the European Commission should be able to adopt common specifications in respect of the requirements for high-risk AI systems, we believe that this should only be the case after a standardisation request has resulted in a standard which is insufficient and does not address specific safety or fundamental right concerns. Therefore, we suggest updating Article 41 to reflect that "Where harmonised standards referred to in Article 40 do not exist, the Commission shall issue a standardisation request in accordance with Article 10 of Regulation 1025/2012 [and] If the Commission considers that the resulting harmonised standards are insufficient and do not address specific safety or fundamental right concerns, the Commission may, […] adopt common specifications."

Enforcement

We are concerned about the possibility for market surveillance authorities to request access to source code as per Article 64.2. Disclosure of source code could seriously put at risk important trade secrets and IP rights, and contravenes established best practices for digital trade at international level. Furthermore, we do not believe the source code of an AI system would be necessary for market surveillance activities. Therefore, we believe that this provision should be deleted.

Also, the Commission's proposal provides for the possibility of a national competent authority in each Member State, with no single point of contact. This framework appears to be inconsistent with the intent of the Regulation, which is to create a harmonised regulation on the use of AI systems in the EU.

Application of the Regulation

Referring to Article 85.2, and with the timelines of the Medical Device regulatory framework in mind, we believe a timeline greater than 24 months will be required to hire and train the necessary workforce in the relevant notified bodies, and to ensure harmonised standards are available.

Conclusion

In conclusion, we thank the Commission again for the opportunity to offer our contribution on the proposed Regulation and welcome the Commission's emphasis for a risk-based approach, regulating specific uses of AI systems and not the AI technology itself.

IBM believes certain aspects of the Regulation could be clarified, especially to better delineate AI tools and AI systems and outline the resulting obligations of the various actors involved in supplying, training, deploying and using AI systems.

We hope that our comments are useful and look forward to continuing our contribution to the debate in the months ahead.