SOCIETE GENERALE

European Commission - Draft regulation establishing harmonized rules on Artificial Intelligence

Position paper / “Have your say”

Société Générale

29 July 2021

CONTENTS

Societe Generale welcomes the opportunity given by the European Commission to comment on its "Draft regulation establishing harmonized rules on Artificial Intelligence".

General comments

The European Commission (EC) has taken into account / integrated the comments and recommendations made, in particular by financial players, in response to the consultation it launched following the publication of its White Paper on Artificial Intelligence (AI) in February 2020.

In view of the present and future challenges linked to the use of AI and the complexity of the subject, the draft EC regulation seems to us to demonstrate a fairly high degree of consistency in its fundamentals, even if, and it is the subject of the comments and observations of this "Have your say", it seems to us that certain subjects require special attention and should be the subject of some adjustments.

Definition of Artificial Intelligence (Article 3)

The text (at level 1) gives a *functional* definition of AI[1] in line with the definitions of AI usually proposed by industry or academics.

The definition proposed by the EC explicitly mentions that AI can produce outputs **"for a given set of human-defined objectives"**, which, by excluding for AI any possibility of intentionality[2], de facto excludes "strong AI" or "artificial consciousness", thus making it possible, and in our view rightly, to consider AI as nothing more than a *technical object*, thereby allowing to rely on a **principle of technological neutrality**.

Influence of an AI system on its environment

We understand from Article 3 that to be qualified as an *AI system*, the software must (i) include one or more AI techniques listed in the appendix and (ii) produce outputs (content, forecasts, recommendations, decisions) that ***influence*** the environment with which they interact. Thus, at the opposite, software comprising an AI technique and producing outputs that do not influence the environment with which they interact would not enter into the definition of an AI system. We therefore understand that to confirm the qualification as an AI system of software based on AI techniques and producing outputs, it is necessary to check whether these influence the environment with which they interact, and that consequently the stakeholders will have to identify the purpose and impact of the software used on their activities and the individuals concerned.

It therefore seems to us that the definition of AI requires clarification in the sense that the "AI" designation applied to a system: (i) should intrinsically **only depend on the technical nature of the system**, (ii) **should not be contingent**, in the sense that it should not depend on the choice made by human, in a given situation at a given time, to use or not the said system to influence the environment with which it interacts.

[1] 'Artificial intelligence 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, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.
[2] Intentionality: the possibility for an entity to set its own goal or objectives.

Artificial Intelligence Scope

However, although it is likely to evolve according to the evolution of technology, the technical annex (Annex 1), to which this definition refers, identifies a set of techniques and approaches of Artificial Intelligence whose list does not seem to us, as of now, to be fully representative of the field of AI usually recognized by experts in this field.

Some AI approaches or techniques are not mentioned. For example, *Multi-Agent Systems*, which represent an important branch of AI, are not listed in Annex 1.

In contrast, certain approaches or techniques are mentioned in Annex 1, while they are not usually considered to be part of the field of AI. This is particularly the case with so-called "*deterministic*[3] and *predictable*[4]" systems, such as *expert systems* (built on knowledge bases and inference rules) or systems based on classical methods, whether *statistical methods* (Bayesian approach or other) or *search and optimization methods* (linear programming, etc.).

A clarification and refocusing on AI systems therefore seem useful to us, which means defining precisely the *characteristics or functional criteria* that base the distinction between "traditional information systems" and "AI systems".

Proposal to revise the definition of AI

As a consequence of these different observations, we suggest:
- to complete the proposed definition (at level 1) as follows: *"artificial intelligence system (AI system) means software,* ***other than a so-called "deterministic and predictable" system****, 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, generate* ***results which, if the human being so decides[5], may be used to produce*** *outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with"*
- to remove from Annex 1 the technologies referred in points b) et c)
- to include in Annex 1 **only illustrative examples** of AI technologies that comply with the level 1 definition.

Risk proportionality approach

The risk-based approach seems particularly relevant to us and consistent with the theoretical and practical fundamentals of financial players (banks and insurance companies).

It should be noted, however, that the widespread application, by companies and in particular banks, of the precautionary principle could lead them to submit, perhaps more than necessary, certain AI

[3] A system is said to be *deterministic* when it makes it possible to identify positively and unambiguously the causal relationships that determine the results it produces
[4] A system is considered to be *predictable* when it unsurprisingly produces the expected results, which implies in particular that its results are not very sensitive to small variations in initial conditions.
[5] As with any *technical object*, an AI system never does anything by itself, it *only does what the human being decides to do with it!*

systems to the “high-risk AI” qualification, thus leading to the implementation of all the induced governance.

Consequently, it would seem useful to us, for a better effectiveness of the qualification “high-risk AI” to restrict the scope of high-risk AI systems, by creating an intermediate category between “high-risk AI” and “low-risk AI”, so as not to consider the “high-risk AI” category as a “catch-all by default” category. To be compared with our remarks made later on the scope of high-risk AI and Annex 3: "Only "secondary" information – such as examples or detailed specifications – that is likely to evolve over time, and useful for understanding the main text, should be included in annexes."

Principle of technological neutrality

Generally speaking, the draft regulation should rely as much as possible on the *principle of technological neutrality* with requires that the use of AI should not increase the requirements (control, governance, transparency, etc.) *per se*. Such requirements should be based on *risk per use case* and not on technology. The draft regulation should therefore be limited to (minimum) requirements to address the specific risks and problems associated with the use of AI, provided that they are not already covered by existing regulations, or that it is not possible to adjust or supplement the existing regulatory framework.

“High-risk AI systems” are defined in Annex 3 through use cases (“intended uses”) listed by domain, and not according to the level of intrinsic risk that could present the different techniques (methods or approaches) of Artificial Intelligence. This approach seems to us to be relevant.

However, the text would be clearer if the term "high-risk AI systems" were systematically replaced by the term "high-risk uses (of AI systems)", which is better suited to refer to what is actually covered by Annex 3, i.e. a set of intended uses. Following the same logic, the terms “unacceptable-risk AI systems”, “low-risk AI systems” and “minimal-risk AI systems” should be systematically replaced by, respectively, the terms “unacceptable-risk uses (of AI systems)”, “low-risk uses (of AI systems)” and “minimal-risk uses (of AI systems).

We therefore wish to emphasize, in a logic guided by the principle of technological neutrality, that in a “risk-based approach”, **it is the intended uses that are at risk**, whatever the technologies used (AI systems or systems other than AI), **and not AI systems**.

In terms of risk, the main question that should be answered by the entire regulatory framework seems to us to be the following: "For each type of intended use, what does the human being (or a human organization) allow himself to do or not to do, and under what conditions, and this whether or not the human uses technology, and if he does, whatever the nature of this technology (AI systems or systems other than AI)?”

High-Risk AI Scope

We question the inclusion of 5(b) in Annex 3: *“AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use”*. We note the concern in Recital 37 that *“AI systems used to evaluate the credit score or creditworthiness of natural persons […] determine those persons’ access to financial resources or essential services […] or may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination.”* However, these risks are present whenever credit scoring is performed, **regardless of the technology used**. Furthermore, regulated

financial services firm would be in breach of existing legislation relating to consumer protection (for example the EBA Guidelines on Loan Origination and Monitoring Section 4.3.4[6]) if they were to ignore or fail to mitigate such risks.

It is therefore not clear what the purpose of including 5(b) in Annex 3 is, unless it is intended to apply only to unregulated firms, in which case this should be made clear.

The definition of the purposes that generate the qualification of high-risk AI seems to us to leave too many possibilities for interpretation. As a result, they could be interpreted differently in different countries of the European Union. In particular, we understand from recital 38 and Annex 3, that with regard to the analysis of large volumes of data allowing the identification of links or behaviours useful for the analysis of criminality involving persons physical, only software used by law enforcement authorities can be classified as high risk. Thus, software used by private actors, the results of which could be used by authorities in charge of enforcing the law, would be excluded from the perimeter of high-risk AI.

We ask for confirmation that software used by private actors, the results of which could be used by authorities in charge of enforcing the law, are excluded from the perimeter of high-risk AI.

Transparency and explainability

The draft regulation mainly refers to the notion of *transparency* and almost never to that of *explainability*, which seems to us to be the right approach to tackle the functioning of AI systems, the notion of transparency being broader and more encompassing than that of explainability[7].

Indeed, it is advisable to be attentive to what covers the concept of explainability, by reminding that if one excludes (as we propose it) from the definition of AI the so-called "deterministic and predictable" systems, then AI systems are never *fully explainable*, in the sense of *white boxes* that would make it possible to identify positively and unambiguously the path followed by an algorithm to produce an output. The challenge is therefore to have AI systems that are, not fully explainable, but *sufficiently transparent* (sufficiently explainable, traceable and auditable), having regard to the rules of the art in terms of AI, to the expectations users and the requirements set by supervisors, taking into account the nature of the use cases and the potential risk levels that AI systems may present.

In addition, in some cases, explainability requirements could pose difficulties when certain models (credit scoring, etc.) are subject to business secrecy.

We would like to remind that the banking and financial sector is already highly regulated with important confidentiality obligations (banking secrecy which falls within the scope of criminal offence, business secrecy, etc.) which aim to protect our customers and partners, which are not necessarily compatible with transparency and explainability requirements. We propose that the elements related to transparency and explainability as well as to technical documentation fall within the competence of the European Central Bank within the framework of its supervision so as to best calibrate the requirements of European regulations in relation to our sector.

[6] https://www.eba.europa.eu/regulation-and-policy/credit-risk/guidelines-on-loan-origination-and-monitoring Paragraph54 *"When using automated models for creditworthiness assessment and credit decision-making, institutions should understand the models used, and their methodology, input data, assumptions, limitations and outputs…"*

[7] According to the HLEG of European Commission : "A crucial component of achieving Trustworthy AI is transparency which encompasses three elements: 1) traceability, 2) explainability and 3) open communication about the limitations of the AI system."

Human oversight

The introduction of the concept of *human oversight* among the obligations for monitoring high-risk systems seems particularly appropriate to help control or reduce the risks of AI systems. This measure directly contributes to the development of trustworthy AI, by encouraging actors, beyond and in addition to regulation, to implement ethical approaches and principles.

Human oversight, by allowing at any time or regularly to verify the correct adequacy (for example in the case of self-learning AI) between the expected behaviour of an AI system and its actual behaviour, allows to maintain a high level of transparency of this system.

It should be noted that for banks, *human oversight* is already required in the model risk governance system. Only the degree of monitoring varies according to the materiality of the model.

Extraterritoriality

The principle that the rules apply to all providers and users of AI systems, including those located outside the EU, when those systems affect people in the EU, seems to us to be perfectly relevant. We will have to see in practice how this principle can be applied.

We are waiting for the European authorities to provide details on this principle.

Supervision

For banks and financial services, we understand that supervision under the AI regulation should be exercised through their traditional supervisors (BCE / ACPR for France), which seems perfectly consistent to us, thus avoiding the multiplication of supervisors for the same subject, potentially a source of confusion and cumbersome for the place.

Regulatory framework and governance system

However, the regulatory framework proposed by the EC seems relatively cumbersome, with a main text accompanied by numerous annexes, several of which are subject to the management of a new agency ("European Artificial Intelligence Board").

The proposed mechanism also seems relatively complex, and probably long, to implement and stabilize, due to the large number of actors involved in its functioning and governance (EC, European AI Board, competent national authorities, etc.).

Thus, the *Union safeguard procedure* (Article 58) seems very cumbersome and complex to implement, and the purpose of this procedure remains difficult to understand.

The safeguard procedure would gain in understanding if the rationale and purpose of this procedure were clarified and simplified. The processing times for the safeguard procedure are very long. These elements are not guaranteeing legal certainty.

Design of the text and its Annexes

In addition, there is a certain imbalance in the text between the relatively heavy governance aspects, marked by a strong desire to provide guidance (declaration of high-risk AI systems, creation of a new Board, supervision of sandboxes by national authorities, triggering by the local authorities of cascading assessment procedures for high-risk systems, etc.) and the part devoted to the development of a AI ecosystem.

The main body of the text (level 1) refers to several annexes that the EC could change over time.

Given the importance of the information contained in these annexes (in particular, in Annex 1 the list of AI approaches or techniques, and in Annex 3 the list of high-risk AI systems) and the impacts on the various economic actors that a change in the content of these annexes could have, it would seem useful to us to provide for **probationary periods**, in the event of modification of these annexes having the effect of reclassifying initially low-risk AI systems or applications.

More generally, it seems unwise to us that information as structuring as the list of *AI approaches or techniques*, or the list of *high-risk AI systems*, be contained in annexes, and not at level 1 of the text. From this point of view, the approach adopted by the text may present some form of legal instability.

Only "secondary" information – such as examples or detailed specifications – that is likely to evolve over time, and useful for the understanding of the main text, should be included in annexes.

This is what we have proposed in particular with regard to Annex 1 (see above): after clarifying in the main body of the text (level 1) the definition of AI by excluding so-called "deterministic and predictable" systems, include in Annex 1 nothing but **illustrative examples** of AI technologies that comply with this level 1 definition.

In the same vein, we note that the European Parliament and the Council of the EU delegate to the EC the updating of these structuring annexes. Indeed, in its conception, the first level text – proposed by the EC for adoption by the Parliament and the Council – delegates the updating of the annexes to the EC without there being any review by the Parliament and the Council.

Public list of high-risk AI systems

Building a public list of high-risk AI systems can be problematic (business secrecy, competition problem, etc.) without bringing in our opinion any real added value.

It would probably be better to replace it with lists internal to each company or organization, made available to national supervisors, similar to what is done today on the GDPR.

Technical documentation

Providing any technical information allowing the best use of AI systems is a good thing. Nevertheless, certain elements are very sensitive, such as the programming code (Annex IV) and can raise issues related to “manufacturing” secrecy, business secrecy, or competition, etc.

We would like to remind that banks and financial institutions are already subject to regulations specific to their sector, as well as to important rules of compliance. We therefore propose that the elements related to technical documentation as well as to transparency and explainability fall within the competence of the European Central Bank within the framework of its supervision in order to best calibrate the requirements of European regulations in our sector.

Relationship with subcontractors / Accountability

Note (Article 28) that in the event of the use of external / third parties (GAFA, etc.) AIs, it is the final distributor who is liable for all the obligations (declaration, accountability) provided for by regulations, and that nothing is charged to subcontractors or suppliers who do not always produce the right information.

For illustration purposes:

- The question of responsibilities in the event of a chain of actors arises from the articulation of Articles 28, 16, 13 and the definition of User, who is necessarily a professional. Article 16 defines the obligations of providers. In particular, there is the management of the system's risk over time, governance to test and validate the system, the establishment of technical documentation, and the information necessary for the transparency of the system, including its purpose, the consequences of possible misuse, and predetermined changes to the system as well as the ability for the system to record its key events ("logs").
- The bank, if it puts AI back on the market or if it integrates it into a service under its name or brand, in turn becomes a provider and recovers the obligations of Article 16. This raises the question of an AI that it would integrate into systems or that it would use without these being placed on the market or marketed, but which would have applications concerning customers. What is the scope of the term "placed on the market" as well as "integrated into a service under its own name or brand"? We use AI from other vendors, which we use for use with our customers, under our own name.
- The obligations of the original Provider also cease if the user has changed the purpose of the AI or has made a substantial change to it.
- In addition, Article 28 specifies that the initial Provider will then no longer be considered as a provider. The question is therefore that of the recourse of the new provider against the old one and the maintenance of the documentation which had fallen to the first provider.

The answers to these questions should be part of the forthcoming European Regulation on the liability regime for AI systems.

The issue of subcontracting and chains of responsibility is not addressed.

It would therefore be desirable, with the aim of consistency and harmonisation (in particular to avoid negotiating contracts on a case-by-case basis), to add to the draft regulation *mandatory information elements* to be included in subcontracts (documentation, audit reports, etc.) that a third-party AI provider should be required to provide to the end-user or distributor to enable them to assess their risks and fulfil their own obligations (reporting, governance, etc.)

Regulatory consistency

It is necessary to ensure a good articulation between the European regulation on AI and other European regulations, in particular that to come on Data, or that to come on AI (the framework of responsibility for the exploitation of AI systems), as well as consistency with national law on the different liability regimes. It is important to avoid the “regulatory thousand-sheet” effect with an accumulation of different texts dealing with the same subjects, both at European and national level, in order to obtain better legibility and application of the rules by the different operators of AI, to allow texts to be harmonized, and to best promote the establishment of “level playing fields”.

Sandboxes

The development of Data within companies requires testing phases to better understand the reasonable expectations of customers.

Thus, participation in AI regulatory sandboxes would find its full utility if it allowed **unit tests** to be carried out (including for high-risk applications), giving the possibility of best approaching **customer expectations**, all of this within a framework that makes it possible to keep compliance requirements and implementation costs at a **reasonable level**.

SOCIETE GENERALE

Commission Européenne - Projet de réglementation établissant des règles harmonisées sur l'Intelligence Artificielle

Position paper / "Have your say"

Société Générale

29 juillet 2021

SOMMAIRE

Société Générale se félicite de l'opportunité donnée par la Commission Européenne de commenter son « Projet de réglementation établissant des règles harmonisées sur l'Intelligence Artificielle ».

Observations générales

La Commission Européenne (CE) a pris en compte / intégré les commentaires et recommandations faits, notamment par les acteurs de la finance, en réponse à la consultation qu'elle avant lancée suite à la publication de son Livre Blanc sur l'Intelligence Artificielle (IA) en février 2020.

Eu égard aux enjeux présents et à venir liés à l'usage de l'IA et à la complexité du sujet, le projet de réglementation de la CE nous semble faire preuve d'une assez grande cohérence dans ses fondamentaux, même si, et c'est l'objet des commentaires et des observations de ce "Have your say", il nous semble que certains sujets nécessitent une attention particulière et devraient faire l'objet de quelques d'ajustements.

Définition de l'Intelligence Artificielle (Article 3)

Le texte (au niveau 1) donne une définition *fonctionnelle* de l'IA[8] en phase avec les définitions de l'IA habituellement proposées ou retenues par les industriels ou les académiques.

La définition proposée par la CE mentionne explicitement que l'IA peut produire des outputs *« pour un ensemble donné d'**objectifs définis par l'homme** »*, ce qui, en excluant pour l'IA toute possibilité d'intentionnalité[9], exclut de fait l'IA dite « forte » ou « conscience artificielle », permettant ainsi, et selon nous à juste titre, de considérer l'IA comme rien d'autre qu'un objet technique, autorisant par là-même de s'appuyer sur un ***principe de neutralité technologique.***

Influence d'un système d'IA sur son environnement

Nous comprenons à la lecture de l'Article 3 que pour être qualifiés de système d'IA, les logiciels devront (i) comprendre une ou plusieurs techniques d'IA listées dans l'annexe et (ii) produire des outputs (contenu, prévisions, recommandations, décisions) qui ***influencent*** l'environnement avec lequel ils interagissent. Ainsi, à contrario, un logiciel comprenant une technique d'IA et produisant des outputs qui n'influencent pas l'environnement avec lequel ils interagissent n'entreraient pas dans la définition d'un système d'IA. Nous comprenons donc que pour confirmer la qualification en système d'IA d'un logiciel basé sur des techniques d'IA et produisant des outputs, il faut vérifier si ces derniers influencent l'environnement avec lequel ils interagissent, et qu'en conséquence les acteurs devront identifier la finalité et l'impact des logiciels utilisés sur leurs activités et les individus concernés.

Il nous semble donc que la définition de l'IA nécessite une clarification en ce sens que la dénomination « IA » appliquée à un système : (i) ne devrait dépendre intrinsèquement **que de la nature technique du système**, (ii) **ne devrait pas être contingente**, en ce sens qu'elle ne devrait pas dépendre du choix fait par l'homme, dans une situation donnée à un moment donné, d'utiliser ou non ledit système pour influencer l'environnement avec lequel il interagit.

[8] On entend par « système d'intelligence artificielle » (système d'IA), un logiciel qui est développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit.

[9] Intentionnalité : la possibilité pour une entité de fixer son propre but ou ses propres objectifs.

Scope de l'Intelligence Artificielle

Cependant, bien qu'elle soit susceptible d'évoluer en fonction de l'évolution de la technologie, l'annexe *technique* (Annexe 1), à laquelle cette définition fait référence, répertorie un ensemble de techniques et d'approches d'Intelligence Artificielle dont la liste ne nous paraît pas, à ce jour, représentative du champ de l'IA habituellement reconnu par les experts de ce domaine :

Certaines approches ou techniques de l'IA ne sont pas mentionnées. Par exemple, les *Systèmes Multi-Agents*, qui représentent une branche importante de l'IA, ne figurent pas en Annexe 1.

A l'opposé, certaines approches ou techniques sont mentionnées en Annexe 1, alors qu'elles ne sont habituellement pas considérées comme faisant partie du domaine de l'IA. C'est le cas notamment des systèmes dits « *déterministes*[10] et *prévisibles*[11] », comme les *systèmes experts* (fondés sur des bases de connaissances et des règles d'inférence) ou les systèmes s'appuyant sur des méthodes classiques, qu'il s'agisse de *méthodes statistiques* (approche bayésienne ou autre) ou de *méthodes de recherche et d'optimisation* (programmation linéaire, etc.).

Une clarification et un recentrage sur les systèmes d'IA nous semblent donc utile, ce qui suppose de définir quels sont précisément les *caractéristiques ou critères fonctionnels* qui fondent la distinction entre les « systèmes d'information traditionnels » et les « systèmes d'IA ».

Proposition de révision de la définition de l'IA (Article 3)

En conséquence de ces différentes observations, nous suggérons :
- de compléter la définition proposée dans le corps principal du texte de la manière suivante : *« on entend par « système d'intelligence artificielle » (système d'IA), un logiciel,* ***autre qu'un système dit « déterministe et prévisible »****, développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'Annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats* ***susceptibles, si l'homme en décide ainsi[12], d'être utilisés pour produire des « effets »*** *tels que des contenus, des prédictions, des recommandations ou des décisions influençant l'environnement avec lequel le système interagit »*
- de **supprimer de l'Annexe 1 les technologies référencées aux points b) et c)**
- de n'inclure dans l'Annexe 1 **que des exemples illustratifs** de technologies d'IA conformes à la définition de niveau 1.

Approche de proportionnalité aux risques

L'approche par les risques nous semble pertinente et en cohérence avec les fondamentaux des acteurs de la finance (banques et assurances).

A noter toutefois que l'application généralisée, par les entreprises et notamment les banques, du principe de précaution pourraient les conduire à soumettre, peut-être plus que nécessaire, certains

[10] Un système est dit *déterministe* lorsqu'il permet d'identifier de manière certaine et univoque les relations de causalité qui déterminent les résultats qu'il produit.
[11] Un système est réputé *prévisible* lorsqu'il produit sans surprise les résultats attendus, ce qui suppose en particulier que ses résultats soient peu sensibles à de faibles variations des conditions initiales.
[12] Comme pour tout *objet technique*, un système d'IA ne fait jamais rien par lui-même, il ne fait *que ce que l'homme décide de faire de lui* !

systèmes d'IA à la qualification « IA à haut risque » entraînant alors la mise en place de toute la gouvernance induite.

En conséquence, il nous semblerait utile, pour une meilleure *efficacité* de la qualification « IA à haut risque » de restreindre le périmètre des systèmes d'IA à haut risque, en créant une catégorie intermédiaire entre « IA à haut risque » et « IA à risque faible », afin de ne pas considérer la catégorie « IA à haut risque » comme une catégorie « attrape-tout par défaut ». A rapprocher de nos remarques faites plus loin sur le périmètre des IA à haut risque et l'Annexe 3 : « Seules des informations « secondaires » – comme des exemples ou des spécifications détaillées – susceptibles d'évoluer dans le temps, et utiles à la compréhension du texte principal, devraient figurer en annexes. »

Principe de neutralité technologique

De manière générale, le projet de réglementation devrait s'appuyer autant que possible sur le *principe de neutralité technologique* qui implique que l'utilisation de l'IA ne devrait pas augmenter les exigences (contrôle, gouvernance, transparence, etc.) *per se*. De telles exigences devraient être fondées sur les risques par cas d'usage et non sur la technologie. Le projet de réglementation devrait donc se limiter aux exigences (minimales) pour faire face aux risques et aux problèmes spécifiquement liés à l'utilisation de l'IA, pour autant qu'ils ne soient pas déjà couverts par les réglementations existantes ou qu'il ne soit pas possible d'ajuster ou de compléter le cadre réglementaire existant.

Les « systèmes d'IA à haut risque » sont définies dans l'Annexe 3 au travers de cas d'usage (« destinations ») répertoriés par domaine, et non pas en fonction du niveau de risque intrinsèque que pourraient présenter les différentes techniques (méthodes ou approches) d'Intelligence Artificielle. Cette approche nous semble pertinente.

Cependant, le texte gagnerait en clarté si l'expression *« systèmes d'IA à haut risque »* était systématiquement remplacée par l'expression *« usages à haut risque (des systèmes d'IA) »*, qui est plus apte à désigner ce qui est effectivement visé par l'Annexe 3, c'est-à-dire un ensemble d'usages. Selon la même logique, les expressions *« systèmes d'IA à risque inacceptable »*, *« systèmes d'IA à risque faible »* et *« systèmes d'IA à risque minimal »* devraient être systématiquement remplacées par, respectivement, les expressions *« usages à risque inacceptable (des systèmes d'IA) »*, *« usages à risque faible (des systèmes d'IA) »* et les *« usages à risque minimal (des systèmes d'IA) »*.

Nous souhaitons ainsi souligner, dans une logique guidée par le principe de neutralité technologique, que dans une « approche par les risques », **ce sont les usages qui sont à risques**, et ce quelles que soient les technologies utilisées (systèmes d'IA ou autres systèmes que l'IA), **et non pas les systèmes d'Intelligence Artificielle**.

En matière de risque, la principale question à laquelle devrait répondre l'ensemble du dispositif réglementaire nous semble être la suivante : « Pour chaque type d'usage, qu'est-ce que l'homme (ou une organisation humaine) s'autorise à faire ou non, et selon quelles conditions, et ceci que l'homme recourt ou non à la technologie, et s'il y recourt, quelle que soit la nature de cette technologie (systèmes d'IA ou autres systèmes que l'IA) ? »

Scope des IA à Haut risque

Nous nous interrogeons sur l'inclusion du 5(b) en Annexe 3 : *« Les systèmes d'IA destinés à être utilisés pour évaluer la solvabilité des personnes physiques ou pour établir leur note de crédit, à l'exception des systèmes d'IA mis en service par de petits fournisseurs et utilisés exclusivement par ces derniers »*. Nous

notons la préoccupation exprimée au considérant 37 selon laquelle *« les systèmes d'IA utilisés pour évaluer la note de crédit ou la solvabilité des personnes physiques […] déterminent l'accès de ces personnes à des ressources financières ou à des services essentiels […] ou qu'ils peuvent conduire à la discrimination à l'égard de personnes ou de groupes et perpétuer des schémas historiques de discrimination »*. Cependant, ces risques sont présents chaque fois que la notation de crédit est effectuée, **quelle que soit la technologie utilisée**. En outre, une entreprise de services financiers réglementée enfreindrait la législation existante relative à la protection des consommateurs (par exemple, les lignes directrices de l'EBA sur l'obtention et le suivi des prêts, section 4.3.4) si elle ignorait ou ne parvenait pas à atténuer ces risques.

L'objectif de l'inclusion du point 5(b) à l'Annexe 3 n'est donc pas clair, à moins qu'il ne soit destiné à s'appliquer uniquement aux entreprises non réglementées, auquel cas il convient de le préciser.

La définition des finalités qui génèrent la qualification d'IA à haut risque nous semble laisser trop de possibilités d'interprétation. De ce fait, elles pourraient être interprétées de façon différente selon les pays de l'Union Européenne. En particulier, nous comprenons au regard du considérant 38 et de l'Annexe 3, qu'en ce qui concerne l'analyse de larges volumes de données permettant l'identification de liens ou comportements utiles à l'analyse de la criminalité concernant des personnes physiques, seuls les logiciels utilisés par des autorités en charge de faire appliquer le droit peuvent être qualifiées de haut risque. Ainsi, des logiciels utilisés par des acteurs privés, dont les résultats pourraient être utilisés par des autorités en charge de faire appliquer le droit, seraient exclus du périmètre des IA à haut risque.

Nous demandons la confirmation que les logiciels utilisés par des acteurs privés, dont les résultats pourraient être utilisés par des autorités en charge de faire appliquer le droit, sont exclus du périmètre des IA à haut risque.

<u>Transparence et explicabilité</u>

Le projet de réglementation fait principalement référence à la notion de *transparence* et presque jamais à celle d'*explicabilité*, ce qui nous semble être la bonne approche pour aborder le fonctionnement des systèmes d'IA, la notion de *transparence* étant plus large et plus *englobante* que celle d'explicabilité[13].

En effet, il convient d'être attentif à ce que recouvre la notion d'explicabilité, en rappelant que si l'on exclut (comme nous le proposons) de la définition de l'IA les systèmes dits « déterministes et prévisibles », alors les systèmes d'IA ne sont jamais *totalement explicables*, au sens de *boîtes blanches* qui permettraient d'identifier de manière certaine et univoque le chemin suivi par un algorithme pour produire un output. L'enjeu est donc bien de disposer de systèmes d'IA qui soient, non pas totalement explicables, mais *suffisamment transparents* (suffisamment explicables, traçables et auditables), eu égard aux règles de l'art en matière d'IA, aux attentes des utilisateurs et aux exigences fixées par les superviseurs, compte tenu de la nature des cas d'usage et des niveaux de risques potentiels que peuvent présenter les systèmes d'IA.

Par ailleurs, dans certains cas, les exigences d'explicabilité pourraient poser des difficultés lorsque certains modèles (credit scoring, etc.) sont soumis au secret des affaires.

[13] According to the HLEG of European Commission : "A crucial component of achieving Trustworthy AI is transparency which encompasses three elements: 1. traceability, 2. explainability and 3. open communication about the limitations of the AI system."

Nous souhaitons rappeler que le secteur bancaire et financier est déjà très réglementé avec des obligations de confidentialité importantes (secret bancaire qui relève de l'infraction pénale, secret des affaires…) qui visent à protéger nos clients et partenaires, qui ne sont pas nécessairement compatibles avec des exigences de transparence et d'explicabilité. Nous proposons que les éléments liés à la transparence et l'explicabilité ainsi qu'à la documentation technique soient du domaine de compétence de la Banque Centrale Européenne dans le cadre de sa supervision de façon à calibrer au mieux les exigences de la réglementation européenne par rapport à notre secteur.

Surveillance humaine

L'introduction de la notion de *surveillance humaine* parmi les obligations de suivi des systèmes à haut risque nous semble appropriée pour contribuer à la maîtrise ou la réduction des risques des systèmes d'IA. Cette mesure contribue directement au développement d'*IA dignes de confiance*, en encourageant les acteurs, au-delà et en complément de la réglementation, à la mise en œuvre d'approches et de principes éthiques.

La *surveillance humaine*, en permettant à tout moment ou régulièrement de vérifier la bonne adéquation (par exemple dans le cas d'IA auto-apprenantes) entre le comportement attendu d'un système d'IA et son comportement effectif, permet de maintenir un haut niveau de transparence de ce système.

A noter que pour les banques, la *surveillance humaine* est déjà requise dans le dispositif de gouvernance des risques de modèle. Seul le degré de surveillance varie en fonction de la matérialité du modèle.

Extra-territorialité

Le principe selon lequel les règles s'appliquent à tous les fournisseurs et utilisateurs de systèmes d'IA, y compris ceux situés en dehors de l'UE, lorsque ces systèmes affectent des personnes dans l'UE, nous semble tout-à-fait pertinent. Il faudra voir dans les faits comment ce principe va pouvoir être appliqué.

Nous attendons que les autorités européennes apportent des précisions sur ce principe.

Supervision

Pour les banques et les services financiers, nous comprenons que la supervision au titre de la réglementation de l'IA devrait s'exercer par l'intermédiaire de leurs superviseurs traditionnels (BCE / ACPR pour la France), ce qui nous semble parfaitement cohérent, permettant ainsi d'éviter la multiplication de superviseurs pour un même sujet, potentiellement source de confusion et de lourdeurs pour la place.

Dispositif réglementaire et système de gouvernance

Le dispositif réglementaire proposé par la CE nous semble cependant relativement lourd, avec un texte principal accompagné de nombreuses annexes dont plusieurs sont soumises à la gestion d'une nouvelle agence (« Conseil européen de l'IA »).

Le dispositif proposé nous semble également relativement complexe, et probablement long, à mettre en œuvre et à stabiliser, de par le grand nombre d'acteurs intervenant dans son fonctionnement et sa gouvernance (CE, Conseil européen de l'IA, autorités nationales compétentes, etc.).

Ainsi, la procédure de *Union safeguard procedure* (Article 58) semble très lourde et complexe à mettre en œuvre, et la finalité de cette procédure reste difficile à comprendre.

La procédure de sauvegarde gagnerait en compréhension si le rationnel et la finalité de cette procédure étaient explicités et simplifiés. Les délais de traitement de la procédure de sauvegarde sont très longs. Ces éléments ne sont des pas des gages de sécurité juridique.

Conception du texte et de ses Annexes

Par ailleurs, on peut noter un certain déséquilibre du texte entre les aspects de gouvernance, relativement lourds, empreints d'une forte volonté d'encadrement (déclaration des systèmes d'IA à haut risque, création d'un nouveau Conseil, supervision des bacs à sable par les autorités nationales, déclenchement par les autorités locales de procédures d'évaluation en cascade pour les systèmes à haut risque, etc.) et la part consacrée au développement d'un écosystème de l'IA.

Le corps principal du texte fait référence à plusieurs annexes que la CE serait susceptible de faire évoluer dans le temps.

Compte-tenu de l'importance des informations contenues dans ces annexes (notamment, en Annexe 1 la liste des *approches ou techniques d'IA*, et en Annexe 3 la liste des *Systèmes d'IA à haut risque*) et des impacts sur les différents acteurs économiques qu'une évolution du contenu de ces annexes pourrait avoir, il nous semblerait utile de prévoir des **périodes probatoires**, en cas de modification de ces annexes ayant pour effet de reclassifier des systèmes ou des applications d'IA initialement à faible risque.

Plus généralement, il nous semble peu judicieux que des informations aussi structurantes que la liste des *approches ou techniques d'IA*, ou la liste des *systèmes d'IA à haut risque*, soient contenues dans des annexes, et non pas au premier niveau dans le corps du texte principal. De ce point de vue, l'approche adoptée par le texte peut présenter une certaine forme d'instabilité juridique.

Seules des informations « secondaires » – comme des exemples ou des spécifications détaillées – susceptibles d'évoluer dans le temps, et utiles à la compréhension du texte principal, devraient figurer en annexes.

C'est ce que nous avons proposé en particulier concernant l'Annexe 1 (voir plus haut) : après avoir clarifié dans le corps principal du texte la définition de l'IA en excluant les systèmes dits « déterministes et prévisibles », n'inclure dans l'Annexe 1 rien d'autre que des **exemples illustratifs** de technologies d'IA conformes à cette définition de niveau 1.

Dans le même ordre d'idées, nous constatons que le Parlement Européen et le Conseil de l'UE délégueraient à la CE la mise à jour de ces annexes structurantes. En effet, dans sa conception, le texte de premier niveau – proposé par la CE pour adoption par le Parlement et le Conseil – délègue la mise à jour des annexes à la CE sans qu'il y ait de revue du Parlement et du Conseil.

Liste publique des systèmes d'IA à haut risque

La constitution d'une liste publique des *systèmes d'IA à haut risque* peut poser problème (secret des affaires, problème de concurrence, etc.) sans apporter à notre sens une réelle valeur ajoutée.

Il serait préférable de la remplacer par des listes internes à chaque entreprise ou organisation, mises à la disposition des superviseurs nationaux, à l'identique de ce qui est fait aujourd'hui sur le RGPD.

Documentation technique

Fournir toute information technique permettant d'utiliser au mieux les systèmes d'IA est une bonne chose. Néanmoins, certains éléments sont très sensibles, comme le code de programmation (Annexe IV) et peut soulever des problèmes liés au secret de fabrication, au secret des affaires, à la concurrence, etc.

Nous souhaitons rappeler que les banques et institutions financières sont déjà soumises à des réglementations propres à leur secteur, ainsi qu'à des règles de déontologie importantes. Nous proposons donc que les éléments liés à la documentation technique ainsi qu'à la transparence et à l'explicabilité soient du domaine de compétence de la Banque centrale européenne dans le cadre de sa supervision de façon à calibrer au mieux les exigences de la réglementation européenne de notre secteur.

Relation avec les sous-traitants / Accountability

A noter (Article 28) qu'en cas d'utilisation d'IA externes/de tiers (GAFA, etc.), c'est le distributeur final qui est redevable de toutes les obligations (de déclaration, d'accountability) prévues par la réglementation, et que rien n'est mis à charge des sous-traitants ou fournisseurs qui pourtant ne produisent pas toujours les bonnes informations.

A titre d'illustration :

- La question des responsabilités en cas de chaîne d'acteurs découle de l'articulation des articles 28, 16, 13 et de la définition de User, qui est nécessairement un professionnel. L'article 16 définit les obligations des Providers. Il y a en particulier la gestion du risque du système dans la durée, la gouvernance destinée à tester et valider le système, l'établissement d'une documentation technique, et des informations nécessaires à la transparence du système, notamment son objectif (« purpose »), les conséquences d'une mauvaise utilisation envisageable, et les changements prédéterminés du système ainsi que la capacité pour le système d'enregistrer ses événements clés (« logs »).
- La banque, si elle replace l'IA sur le marché ou si elle l'intègre à un service sous son nom ou sous sa marque, devient à son tour provider et récupère les obligations de l'article 16. Ceci pose la question d'une IA qu'elle intégrerait à des systèmes ou qu'elle utiliserait sans que ceux-ci soient placés sur le marché ou commercialisés, mais qui auraient des applications concernant des clients. Quelle est la portée du terme « placé sur le marché » ainsi que de « intègre dans un service sous son propre nom ou sa marque » ? Nous utilisons en effet des IA d'autres fournisseurs, que nous utilisons pour des utilisations auprès de nos clients, sous notre propre nom.
- Les obligations du Provider initial cessent aussi si l'utilisateur a modifié l'objectif de l'IA ou en a fait une modification substantielle.

- Par ailleurs l'article 28 précise que le Provider initial ne sera alors plus considéré comme un provider. La question est donc celle du recours du nouveau provider contre l'ancien et du maintien de la documentation qui avait incombé au premier provider.

Les réponses à ces questionnements devraient faire partie du Règlement européen à venir relatif au régime de responsabilité pour l'exploitation des systèmes d'IA.

La question de la sous-traitance et des chaînes de responsabilité n'est pas abordée.

Il serait donc souhaitable, dans un objectif de cohérence et d'harmonisation (pour éviter notamment la négociation de contrats au cas par cas), d'ajouter au projet de réglementation des *éléments obligatoires d'information* à insérer dans les contrats de sous-traitance (documentation, comptes-rendus d'audit, etc.) qu'un tiers fournisseur d'IA devrait être tenu de fournir à l'utilisateur ou au distributeur final pour lui permettre d'évaluer ses risques et de remplir ses propres obligations (de déclaration, de gouvernance, etc.)

Cohérence réglementaire

Il est nécessaire de veiller à une bonne articulation entre la réglementation européenne sur l'IA et les autres réglementations européennes, notamment celle à venir sur la Data, ou celle à venir sur l'IA (le cadre de la responsabilité pour l'exploitation des systèmes d'IA), ainsi qu'une cohérence avec le droit national sur les différents régimes de responsabilité. Il est important d'éviter l'effet de « mille-feuille réglementaire » avec une accumulation de différents textes traitants des mêmes sujets, tant au niveau européen que national, afin d'obtenir une meilleure lisibilité et application des règles par les différents opérateurs de l'IA, de permettre une harmonisation des textes, et de favoriser au mieux la mise en place de conditions de concurrence équitables.

Bacs à sable

La valorisation de la Data au sein des entreprises nécessite des phases de tests permettant d'appréhender au mieux les attentes raisonnables des clients.

Ainsi la participation à des bacs à sable réglementaires de l'IA trouverait sa pleine utilité si elle permettait de réaliser des **tests unitaires** (y compris pour des applications à haut risques) donnant la possibilité d'approcher au mieux les **attentes des clients**, le tout dans un cadre permettant de contenir à un **niveau raisonnable** les exigences de conformité et les coûts de mise en place.