Consultation on the EU AI Act Proposal

Google's submission
July 15, 2021

Table of Contents:

Overview

Google welcomes the publication of the European Commission's proposal for the EU Artificial Intelligence Act (AIA). As we said in our submissions to the whitepaper consultation and inception impact assessment, Google believes that AI is too important not to regulate. The AIA's risk-based, proportional approach strikes an appropriate balance between protecting citizens and managing the risk of high-risk AI applications while supporting innovation and the development of transformative applications of AI technologies. We support the AIA's strong controls for biometric identification systems, a class of technologies that carries unique risks, and which Google has approached with great care. The notion of a need for transparency for certain non-high-risk systems, for example systems that might lead a user to believe they are interacting with a human being, is also important, and consistent with Google's own approach in products like Duplex. Although some fine-tuning and clarification is needed, overall the AIA is helpful in providing a clear legal framework for a trustworthy, risk-based approach to AI in Europe.

This document outlines Google's feedback on the draft regulation, including elements of the proposal that we believe are appropriately crafted and scoped, and key questions and areas of concern that could be refined in subsequent discussions with the co-regulators. Our questions and concerns center around three main themes:

1. *Reflect the complexity of the AI ecosystem in the balance of obligations for different stakeholders*. It will seldom be feasible or effective for providers of general-use AI systems to manage all of the risks associated with potential application in high-risk systems, as is currently envisaged in the AIA. For example, the provider will often not have access to the operational data necessary for post-market monitoring if the AI system has been put into operation by another entity. To address this, we recommend a new class of "deployers" be added to the AIA with responsibility for complying with regulatory requirements associated with deploying general-use AI systems in high-risk applications.

2. *Hold providers, deployers and users to feasible standards*. As currently phrased, certain requirements of the regulation will be extremely difficult or impossible to meet in practice (e.g., the Art 10(3) requirement that datasets be "*free of errors and complete*" demands a level of perfection that is not technically feasible). We do not disagree with the spirit of the requirements, but they should be composed in a fashion that reflects feasible, best practice standards.

3. *Clarify certain provisions to provide legal certainty around scope and protect privacy.* The AIA includes a number of terms and provisions that would benefit from further

clarification to ensure that providers of AI systems understand how the scope and requirements of the AIA apply to their products. For example, the definitions of “safety components” and “significant changes” could be further clarified to provide legal certainty around when systems come in scope of requirements for high-risk AI systems.

A more detailed explanation of these points is offered below, and Google would be happy to answer any additional questions the Commission may have regarding our feedback. We also stand ready to continue engaging constructively with EU co-regulators in the coming months to ensure a well balanced, proportionate, risk-based AI regulation in Europe.

1. Reflect the complexity of the AI ecosystem in the balance of obligations for different stakeholders

The Commission’s proposal is well balanced in principle, providing strong protections against the risks posed by high-risk AI applications (as clearly and concretely specified in Title III), while protecting innovation and competition in lower-risk domains. However, in a practical setting, there are areas that require refinement to appropriately reflect the complexity of the AI ecosystem in assigning responsibilities for ex-ante and ex-post compliance.

The AI ecosystem and value chain are highly complex. Many real-world products contain multiple models, developed by different research and engineering teams, each utilizing multiple datasets from multiple sources, co-developed into a final product by multiple organizations, and deployed by a third party, sometimes with multiple possible use cases. The AIA needs adjustment to ensure that compliance obligations are assigned to the entities best positioned to meet them and undertake meaningful risk management, and to provide greater legal clarity over the balance of obligations for different entities in the value chain.

More clarity is also needed over the point at which AI systems transition from research into operational usage and become potentially in scope of the AIA regulation. Typically, the development of an AI system is a complex, iterative process of experimentation, research, model training and retraining, testing and validation, and redevelopment, often starting without a full set of specific applications in mind, and so it is not always possible to know at the outset how likely it is to fall within the high risk scope.

To address this complexity we would recommend the following revisions:

- Add a class of “deployers” who are best positioned to meet compliance obligations under the AIA, particularly for general-use AI systems used in high-risk applications

- Clarify the definition of a provider in art.3 to focus obligations on providers who deploy products that directly impact people (rather than AI research)
- Make clear that open-source software is out of scope
- Clarify 'due diligence' obligations of providers when using third party datasets

Add a class of "deployers" who are best positioned to meet compliance obligations under the AIA, particularly for general-use AI systems used in high-risk applications

Currently, the obligation for complying with the AIA and undergoing conformity assessment rests with "providers" of high-risk AI systems — defined in Article 3(2) as "*a natural or legal person, public authority, agency or other body that develops an AI system or has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark.*" This allocation of responsibility implicitly assumes that the developer of an AI system will be the one deploying it in a high-risk application, or will be developing the system directly on behalf of a deployer. However this is often not true, especially when it comes to general purpose 'off the shelf' AI tools [such as application programming interfaces (APIs) and open source systems (OSS)] which do not require customisation.

Article 28 goes some way to acknowledging that there will be times when entities deploying and using an AI system must take over certain obligations of "providers" for compliance, such as when they modify the system. Article 29 also acknowledges certain requirements that must be met by users even when they do not meet the definition of a "provider," for example logging requirements. But they do not go far enough. In particular, Article 28 excludes instances where a general purpose product or tool is used without modification. This means that, as currently worded, the responsibility for compliance in such situations would rest with the original provider even though they have little to no visibility or control over how it is being used (including whether it falls into the high-risk or low-risk use case category), and lack the operational access to fulfill many of the requirements, including requirements that are not allocated to the user under Article 29 (e.g., for ensuring appropriate human oversight).

For example, consider a deepfake detector API which could be extremely beneficial to small or medium enterprises (SMEs) seeking to combat manipulated media. In itself, this is not inherently high-risk under the AIA. But such a system might be used by law enforcement, with or without the knowledge or consent of the provider releasing the non-high-risk API, which would constitute a high-risk application of the technology (see Article 6(2) in conjunction with Annex III, Section 6(3)).

Requiring entities developing general use AI systems to comply with the AIA and undergo conformity assessment, simply because there might be the potential for others to deploy them in a high risk setting, could end up capturing a broad swath of systems across a range of maturities that are not specifically intended for high-risk use cases, hampering collaboration

and innovation. Conversely, placing no requirements on "users" or deployers of general use AI systems in high-risk applications (as in the current proposal, unless a user has made modifications) creates a loophole that could heighten risks.

One possible solution might be for companies like Google to add guidelines in product documentation instructing customers not to use general-purpose AI systems for high-risk use cases. This would indicate that these systems are not being "placed on the market" for use in high-risk AI systems. Then, under Article 28(1)(b) - which shifts obligations from the provider to the distributor, importer, user or other third-party if they "modify the intended purpose of a high-risk AI system already placed on the market or put into service" - the use of the system in high-risk applications might represent a modification of the developer's stated intended purpose, and the responsibility would shift accordingly. However, as currently written, it is not clear whether the intended purpose is deemed to be modified under these circumstances.

Furthermore, due to Article 9(2)(b), which requires providers to account for "reasonably foreseeable misuse," even if high-risk uses were explicitly barred by the terms of service, it is unclear whether such a use might constitute a "reasonably foreseeable misuse" of the system. We would urge that an explicit provision be added to clarify this. Otherwise, companies will be forced to take a conservative position, imposing a significant chilling effect on the release of general-use APIs and OSS until the issue is resolved in the courts.

However, in practice this solution is not ideal, as it would incentivise restrictions against using general purpose AI systems in high risk settings, which will undoubtedly block some beneficial use cases. A better solution would be to add a class of "deployers" to the existing categories of "providers," "importers," "distributors," and "users" of AI systems in the AIA. A "deployer" can be defined as "*an entity that puts into service an AI system developed by another entity without substantial modification.*" In practice, this would apply to many AI applications deployed by third parties utilizing open-source software or general purpose APIs. This would allow for the use of general-purpose tools in high-risk applications under appropriate safeguards put in place by the "deployer", rather than restricting the use of these tools or imposing compliance obligations on general-use systems that might deter companies from releasing new technologies for low-risk use cases.

More generally, even in cases where an application is provided by a provider directly to the deployer, and no modifications are made, deployers will often be best positioned to understand downstream use cases and their attendant risks, implement effective risk management strategies, and conduct post-market monitoring and logging, which providers of general use systems are not equipped to do (see Table 1).

This is not to say that providers of AI systems that are deployed by third parties do not have an important role to play in facilitating compliance with the AIA. Deployers would depend on

certain information and capabilities from developers (for example, documentation of how models were trained or mechanisms for human oversight) to comply with some provisions of the AIA. However, deployers are best positioned to understand which controls and risk mitigations are most appropriate to their specific use case, and to implement them appropriately. Given this, the Commission may like to consider whether responsibilities for compliance should be moved to "deployers" in all circumstances, not only involving general purpose systems.

Table 1: Compliance obligations that will be difficult or impossible for providers of general-purpose AI systems to meet under most circumstances

Employ appropriate data governance practices (Article 10): While developers of AI systems often manage the data on which the system is initially trained, many systems ingest data from users as part of their operations, and whether and how that data is retained, used and deleted is often controlled by the deployer of a system. As an example, an entity that develops a system to analyze health records often will not have access to the health records of patients in a hospital that deploys the system, nor will they have any control of how the health system ingests, uses, retains or deletes that data in the course of operating the system. The deployer of the system, in this case the hospital, will be the one to control and process patient data, understand how that data is being used and how patient needs may evolve throughout the life of the product. Note that this challenge is reflected in the proposal already, with Article 29 assigning responsibilities for certain downstream data requirements to the user/deployer rather than the original provider.
Provide technical documentation and keep it up to date throughout the life of the system (Article 11): Technical documentation will require information from the provider of AI systems regarding the training, development, and performance of the system, but information on how the system interacts with other systems used by the deployer, whether and how it is patched and updated, and how the system interacts with the deployer's real-world user data will need to be supplied by the deployer.
Provide transparency to users (Article 13): The provider of an AI system deployed by a third party through a website or app often will not control the actual user interface through which users interact with the system. While documentation about the development, training, and performance of the system may be offered by the provider, the deployer of the system must be responsible for ensuring that relevant information, for example, about application-specific decision-making or opportunities for appeal, is appropriately surfaced to users.
Ensure appropriate human oversight of the system (Article 14): In many, if not most, cases, appropriate human oversight of AI systems will require appropriate training, oversight, and accountability for users of the system. While providers can create mechanisms for incorporating human input and feedback and exercising oversight, ensuring that users are appropriately trained to use the system—and that oversight and accountability structures are in place to ensure the system is used as intended—will necessarily be determined by the deployer of the system.
Ensure that systems are used in such a way that they achieve appropriate levels of accuracy, robustness, and cybersecurity (Article 15): Accuracy, robustness and security are highly dependent on choices made by deployers of AI systems, in addition to the developer of the AI system itself. As an example, deployers will generally determine access control for AI systems they use, conduct network monitoring and threat intelligence activities to identify potential cyber threats, and conduct user training on how to avoid security breaches.

Conduct post-market monitoring of the system's performance (Article 61) and correct for unfair bias: Providers of AI systems often lack direct access to the system as deployed by their customers, meaning that they are unable to monitor system performance, identify indicators of bias, and take steps to correct them. In some cases, deployers may need to work with providers to correct for bias that arises due to features of the system, but in other cases they may be more easily corrected by adjusting any additional data the system is trained on or how the system is used.
Put in place an appropriate risk management system (Article 9): All of these elements are critical to establishing an effective risk management system for AI systems. While developers will need to take effective measures to manage risks associated with the design of the system, only deployers are in a position to evaluate whether mitigations put in place by developers are appropriate to their use case and organization.

Clarify the definition of "provider" to focus on products rather than research

Article 3(2) defines "providers" of high-risk AI systems as "*a natural or legal person, public authority, agency or other body that develops an AI system or has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark.*" In practice, developing an AI system is a complex, iterative process of experimentation, research, model training and retraining, testing and validation, and redevelopment, often starting without a specific application in mind. The point at which this becomes development "with a view to" placing a product on the market is not always clear. Researchers cannot anticipate all potential downstream uses of their research, and even if there is a product idea underpinning development, the final version will often differ substantially from its original vision.

Furthermore, AI research often results in work products like research papers or demos that are not in themselves AI systems, but provide detailed explanations of how such systems can be developed and deployed. Such outputs are essential to driving scientific progress in the field, and the AIA should not discourage their development or sharing by imposing administrative hurdles such as would happen were research publications to be considered as "placing on the market" or "putting into service".

The inclusion of "with a view to" in the definition of "provider" therefore creates uncertainty around general AI research and development activity. Removing this language would clarify that entities should undergo conformity assessment before an AI system is placed on the market or put into service, rather than in the course of general AI research or when publishing a paper.

Make clear that open-source software is out of scope

As written, it is unclear whether releasing open-source software (OSS) could constitute "placing it on the market" or "putting into service" or developing a system "with a view to" it being placed on the market from the point of view of the AIA. This matters because OSS is

critically important to AI innovation, and if the AIA were seen as imposing conformity assessment requirements on OSS it would have a chilling effect on open collaboration in the AI ecosystem.

This problem can be avoided by providing legal clarity that the AIA requirements only come into play for operational AI systems, and thus the onus is on the downstream providers or deployers who have opted to deploy OSS in operational systems to satisfy the obligations.

Clarify 'due diligence' obligations of providers when using third party datasets

Many AI systems are developed using datasets supplied by third parties, whether they are made publicly available or used under license or contract. Guidance would be helpful as to what constitutes sufficient due diligence in assessing these datasets, as part of carrying out a conformity assessment for a high risk AI system. In particular, to what extent are providers required to reevaluate the data from scratch, or can they rely on the dataset creator's representations about the provenance of the data? Requiring each provider that uses a publicly available dataset to independently reassess the dataset from scratch would create significant redundant compliance obligations and add little value.

There will also be instances where there is limited information on a legacy dataset's provenance, because it was not recorded at the time of its creation. Indeed, many of the richest and most widely used datasets in AI development lack such documentation, for example ImageNet or MSCOCO. It would be advisable for there to be a 'grandfather clause' akin to Article 83 to exempt legacy datasets from requirements to provide information about their creation that is no longer possible to determine.

2. Hold providers and deployers to feasible standards

In spirit, the obligations set out under the AIA regarding documentation and transparency, risk and quality management, data governance, accuracy, robustness and security, and human oversight are reasonable, proportional and appropriate. Google's responsible AI practices and AI Principles follow a similar approach. However, some requirements, as written, may be difficult or impossible to comply with in real-world situations. Language should be added to clarify that providers and deployers of high-risk AI systems should take reasonable measures to address risks, consistent with industry best practices, recognizing that there are limits to what is possible with the current state of technology. In particular:

Reflect practical realities of datasets in Article 10

Data governance is an important aspect of AI system development and application, and Article 10 provides helpful directional guidance. However there are several areas that could be improved to better reflect practical challenges of working with datasets and to avoid unintended impacts to privacy mitigation strategies. Specifically:

- Article 10(3) requires that "*training, validation and testing data sets shall be relevant, representative, free of errors and complete*." Real-world datasets will almost never be "free of errors," particularly the large datasets, often including millions or billions of individual data points, used in the most advanced AI applications available today. Furthermore, what constitutes "relevant" and "representative" is often unclear, and there are few standards and metrics to measure them or frameworks to consistently apply them. "Completeness" is also a complex concept for datasets - there will always be additional datapoints that can improve a dataset, but at some point a decision must be made that it is good enough. This article should be modified to require that developers and deployers of AI systems "*take appropriate measures to ensure that training, validation and testing data sets are relevant, representative, and free of errors ~~and complete~~, consistent with industry best practices*."

- Article 10(5) provides helpful clarity around the use of personal data for bias monitoring. In a similar vein, it would be useful to add a GDPR exception for static datasets used in system validation, benchmarking, and bias testing. Static datasets are valuable because they allow developers to see how changes in AI systems over time (whether the result of learned behavior or manual modification) impact performance. Static datasets are also important when comparing the performance of different AI systems (including between providers) as using them removes one source of variability. However currently GDPR data retention limits can make it difficult to maintain static benchmark datasets over time, for example when individual datapoints have to be deleted due to data minimization requirements, changing the composition of the dataset. Providing an explicit exception allowing developers to maintain static datasets for these purposes would be a helpful clarification that would support progress in improving AI system robustness and addressing bias.

- Some differential privacy techniques intentionally introduce noise into datasets in order to prevent the unintentional disclosure of sensitive data. This noise introduces more "error" in the data, and could therefore be interpreted to violate Article 10(3) of the AIA. However, these techniques provide important advantages in protecting sensitive data, and are a critical tool for AI developers to protect citizens' privacy. Article 10 should

include an explicit exception for the use of privacy enhancing technologies such as differential privacy techniques that introduce noise into datasets to preserve user privacy and protect sensitive data.

Human understanding of AI systems should enable effective oversight

Article 14(4)(a) requires that individuals that exercise human oversight of AI systems "*fully understand the capacities and limitations of the high-risk AI system.*" For many AI systems, whether highly complex models with millions or billions of parameters or relatively simple hand-coded models, "fully understanding" the system is effectively impossible. Rather individuals should be required to "adequately understand" the system to exercise effective oversight. Furthermore, while this requirement will necessitate appropriate documentation from the developer of the system, as noted above, the deployer, rather than the developer of the system, will need to ensure that the individual performing oversight has the appropriate understanding.

Align with industry best practices on robustness

Article 15(3) requires that high-risk AI systems be "*resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates.*" AI systems, like any computer or human system, are never 100% free of errors and faults, and what constitutes an appropriate level of resilience is often unclear. To provide legal clarity around expectations, providers and deployers should be required to "*take appropriate technical and organisational measures to ensure that high-risk AI systems are resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, consistent with industry best practices.*"

3. Clarify certain provisions to provide legal certainty around scope and protect privacy

Some definitions and obligations under the AIA would benefit from additional clarification of how they apply in specific situations, and what specifically they require.

Define "subliminal technique" and "materially distort a person's behaviour"

Article 5(1)(a) prohibits "*the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm.*" It is unclear, however, what constitutes a "subliminal

technique" or "materially distorting" a person's behavior. For example, many UX features are meant to influence user behavior, often in subtle ways that may not be immediately obvious to the user. Expanding the language in Recital 16 to further illustrate what is meant by these terms, or including clear definitions in Article 3 would help clarify what systems are in scope of this prohibition.

Clarify the definition of a "safety component"

Article 3(14) defines a "*safety component of a product or system*" as "*a component of a product or of a system which fulfills a safety function for that product or system or the failure or malfunctioning of which endangers the health and safety of persons or property.*" It is unclear whether a "safety function" would be interpreted to include a wide range of features that may be tangentially related to safety or security but not actually safety-critical for the system, for example software used in a radio in a Nest product, or a security feature in a non-safety-critical component of a system.

To remove this legal uncertainty and focus appropriately on real risks to health and safety, we would suggest the following rewording: "*a component of a product or of a system ~~which fulfils a safety function for that product or system or~~ the failure or malfunctioning of which endangers the health and safety of persons or property*."

Protect use of privacy-enhancing techniques

Certain systems make use of machine learning on decentralized datasets (for example, data that is stored on devices) rather than uploading data to the cloud for operational reasons and to help protect user privacy. Federated learning (FL), for example, is a technique used by AI developers to train ML models without centralized data collection, enabling systems to learn and adapt over time from real-world data without collecting user data in centralized datasets. These systems, by design, do not log raw user data to a central server, meaning it may not always be possible for these systems to demonstrate compliance with dataset requirements under Article 10, generate centralized logs as outlined in Article 12, or provide direct access to datasets per Article 64. Without an exception, the AIA would effectively ban the use of FL and other decentralized learning techniques in high-risk systems, undermining opportunities to improve machine learning systems while protecting user privacy. These articles should include an explicit exception for systems that use FL or other on-device learning techniques.

Remove requirements for dataset and source code disclosure

These requirements are overly broad, create unnecessary risks and should be removed. The data governance requirements outlined in Article 10 provide reasonable protection, and giving market surveillance authorities access to datasets themselves would in many cases be

unworkable. Sharing source code is also unwarranted as alternative approaches are available that would be more effective and not undermine trade secrets or security. More specifically:

- Article 64(1) states that "*market surveillance authorities shall be granted full access to the training, validation and testing datasets used by the provider.*" This provision is unworkable on multiple levels and should be struck. In some cases the data will simply not exist. For example, data used to train on-device models is only stored for a limited period on devices, and is not collated centrally. There are also likely to be legacy products developed using datasets that were never retained. In addition, there are serious privacy concerns regarding any sharing of datasets that contain personally identifiable information (PII). This includes the heightened exposure risk inherent in data being retained when it would otherwise have been deleted in line with GDPR's data minimisation principles, and the danger that the data transfer mechanisms enabling it are exploited by malicious actors. More generally, the requirement to grant access appears to contradict the protections offered for training data and the schema by which it is organized by the EU Trade Secrets and Database Directives.

- Article 64(2) says that "*where necessary to assess the conformity of the high-risk AI system... upon a reasoned request, the market surveillance authorities shall be granted access to the source code of the AI system*." But there will always be better methods for verifying the performance of an AI system (ex: input/output auditing) than direct access to source code, rendering this provision superfluous. Furthermore, source code is protected by the EU trade secrets directive, and the confidentiality provisions of Article 70 are insufficient to protect providers' intellectual property rights. A better alternative would be to replace Article 64(2) with something like "*Where necessary to assess the conformity of the high-risk AI system with the requirements set out in Title III, Chapter 2 and upon a reasoned request, AI providers or deployers should support and equip market surveillance authorities with the necessary facilities to carry out robust testing (e.g., input/output auditing) to confirm compliance*."

Define "significant changes" to grandfathered systems

Article 83(2) states that AI systems that are already on the market are exempt from compliance with the AIA and undergoing conformity assessment unless "*those systems are subject to significant changes in their design or intended purpose*." Additional clarification around the definition of a "significant change" is necessary for providers of grandfathered high-risk AI systems to understand when these systems would become subject to the requirements of the AIA and be required to undergo conformity assessment. One option might be to change "significant changes" to "substantial modifications," as used in Article 3(23), to align with existing product regulation as outlined in Recital 66.

Google thanks the Commission for the opportunity to provide our feedback on the draft AIA, and would welcome the opportunity to further discuss our feedback and answer any questions.

[End]