Broadcom's addition feedback on European Commission AI White Paper consultation

Introduction

Broadcom would like to thank the European Commission for launching this public consultation and for presenting the White Paper. We believe that the White Paper is moving in the right direction and is discussing the right questions.

We believe we have a unique perspective to bring in the AI discussions because of the diversity of businesses Broadcom is currently engaged. Broadcom is known for semiconductors manufacturing that is AI-enabled. In addition, Broadcom is using and building AI capabilities in several of its software businesses, for purposes such as information technology (IT) infrastructure management, in support of mainframe technology mostly for financial services, in cybersecurity to detect, prevent and mitigate cyberattacks as well as in payment security to detect and prevent financial fraud. Our unique perspective comes from the different use cases and benefits AI brings as well as the feedback of our customers in using and deploying our AI technology.

Broadcom believes that an EU framework on AI can bring important benefits to the market. It can function as a competitive differentiator and it can set qualitative standards. It can also address liability and risk adoption concerns. In order to achieve that the framework needs to be aligned with existing legislation such as GDPR. It needs to take into account the development and use of AI not just in the EU but especially in other developed markets (US and UK). For instance, the UK ICO has developed a detailed AI auditing framework, while the US Federal Reserve has issued governance models on the use of AI for fraud prevention for several years. Ideally the Commission can identify areas of interoperability between its work and those of countries that are critical in the global economy, while improving on what is already available, by providing a clearer framework that is less susceptible to interpretation. Such a framework needs to facilitate access across the single market and enable European companies to export out of Europe globally. Finally, it needs to maintain a reduced administrative burden both in order to make the EU an appealing AI investment destination and in order to limit the governance burden on SMEs wanting to rely or invest in AI.

AI explainability vs transparency

Whereas transparency may be a desirable end-goal for some uses of AI we believe that for other use cases a suitable level of explainability is a much more desirable property. Full transparency of an AI model enables hostile actors to perform attacks against it or to identify techniques to circumvent it or even in some cases extract data that were used to train it. In use cases such as cybersecurity, fraud prevention or IT infrastructure management explainability is a much more desirable property. It would require that in the cases that AI is offered as a service the provider is able to explain a decision of the AI. This is already a property that is understood in the market and a requirement that customers often put forward. Therefore, it would be something that is in line with existing best practices, provides a suitable level of quality assurance without creating a security risk. Article 22 (2) and (3) of the GDPR regulate automated decision in a similar fashion. Apart from valid legal bases (contract, regulatory requirements, consent), there must be demonstrable safeguards built in, and there must be a correction mechanism (human intervention) if the algorithm reaches a questionable outcome. Recital 71 describes this in the following terms: "*In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view,* ***to obtain an explanation of the decision reached after such assessment and to challenge the decision.***" Making a food safety analogy instead of a privacy one, it's a bit like saying: "I don't need to know the recipe to your magic sauce, but I need assurances that it's not toxic, and I want to be able to add salt and pepper if necessary. I also need a gluten-free version, could you consider creating one?". Finally, it is important to remember that apart from considerations like cybersecurity and fraud there is also a lot of innovation in the form of intellectual property and trade secrets that goes into AI development and training and constitutes a competitive advantage. Whereas transparency

and explainability objectives remain important they should be balanced with the interests of protecting commercial secrets and the investments companies do to effectively compete in a global market.

To what degree are AI decisions/actions explainable/auditable?

The answer to this question depends on the training of AI and the capabilities the technology has to be self-evolving or not. It also depends on the degree of explainability one is expecting as well as the degree of underlying administrative burden/guidance one wants to impose. An AI technology that is made available trained out of the box and is not self-evolving should be explainable within the parameters the manufacturer has set. An AI technology that is self-evolving and has been trained with customer data to which the provider/manufacturer has no access to it becomes progressively difficult to explain by the manufacturer and closer to the customer/user. In a scenario whereby the AI is available to a user/customer as a service the provider will make available aspects of the AI logic to allow the customer to further influence the AI for its specific use cases. For example, it is possible to introduce customer-specific business rules on how AI operates. This can result in disregarding AI findings within certain ranges and thus changing the AI outcomes the customer sees introducing further complexities in explainability. As the AI is getting trained and evolves using customer-provided data and changes to its logic, any explainability requirements should be at the level of the customer/user. To focus and explain a specific decision it becomes necessary to have more detail in the evolution of a model. The greater detail on explainability is expected the greater the requirements for documentation and data recording. Notwithstanding the challenge of keeping such records an additional problem could be the recording of data that is not necessary for business purposes but may be relevant for the behaviour of AI. Auditability is especially hard. Perhaps a more tractable, complementary goal could be reproducibility. An AI may be explainable and even auditable without necessarily being reproducible. Ability to reproduce results is often a key property for analysis, forensic investigation, certification and risk assessment.

How could explainability look like?

Whereas there isn't one universal way to explain AI there is a common desire for explanations to help assess the uncertainty, bias, risk and behaviour of an AI's predictions and actions. Rather than focusing on which data attribute causes a change in the behaviour of the AI due to algorithm "xyz" it would be more effective to focus on questions like:

- Under what circumstances will this AI change its prediction?
- How will this AI perform on never-before-seen inputs?
- How does this AI adapt to new data?
- What are the domains of its decisions and actions, e.g., what can it affect? For example, a smart thermostat that can make decisions about the temperature in all rooms of a dwelling and interacts with a heating system.
- What are the ranges of its decisions and actions, e.g., what can these decisions and actions include? For example, a smart thermostat that can decide to increase the temperature of a radiant heater by up to 5 degrees.
- How does the system evolve? For example, a smart thermostat that learns to improve its temperature prediction and the prediction of a user's comfort range.
- What are the sources and provenance of the data used for training and evolution (which can include retraining)? For example, a smart thermostat that uses historical measurements from the dwelling it serves, as well as historical weather data and measurements from all other dwellings, updated daily, plus outputs from other AI that predicts when a user will arrive home, etc.
- What other AI systems does this depend on? For example, a smart thermostat that uses
    - a local weather prediction service to determine the peak temperature at noon of 32 degrees Celsius will require a change at 7:00 am to cool a space to 24 degrees Celsius by noon
    - AI location services that predict the homeowner will eat lunch at noon today at home

- Does this AI produce randomized outputs or accept random inputs? For example, a smart thermostat that can randomly generate target set points to induce a user to adjust the settings so that the user's preferred temperature ranges by season can be learned and tracked.
- What is the feasible domain of inputs this was trained on, and what novel inputs could it face? For example, a smart thermostat developed for coastal Greek homes that is installed in Finland station and connected to a natural gas heating system instead of solar panels.

Such questions would enable us to understand better the parameters on which the AI operates and thus produce an explanation of its decisions, actions. It enables us to "deconstruct" and "interpret" the decision-making and data feeds of the AI rather than looking for the critical data set that may or may not have made the difference.

What is the role and importance of human expertise and of human oversight?

The human expertise is absolutely quintessential in the development of AI. This is not the case just for the obvious bit of AI developers and data scientist. In addition, there is a clear requirement for domain expertise in the development, monitoring and approval of the models for the specific use case these models would address. For example, in the case of financial fraud, or in the management of payment systems it has been proven that such expertise is required for the development of an effective AI system. This dependency is critical to be recognized because the existence or not of such expertise will likely determine the efficacy of the AI and thus it's reliability.

Human oversight on the development of the AI algorithm is very important because it is that oversight that will provide the context in the decision-making process and help distinguish the right decision in the right context, versus the wrong decision. As algorithms get tested the results are often presented as recommendations enabling humans to provide that context that will enable the right decision. Only after repeated successful use and testing with real-life data can algorithms be trusted and automation on results established. This is a decision that involves a provider as much as the user of AI. The training of AI models based on customer data is a joint investment of the provider and the customer. The process begins with the provider running some tests on customer data to calibrate the AI model and identify the weaknesses, but then evolves with the real-life data of the customer, who ultimately makes the decision that the AI meets its needs.

Where is the biggest AI risk?

There are a lot of scenarios of risk related to actions or omissions by an AI system. These can range from personal data, to cybersecurity, to financial risk, to human right violation such as discrimination. Our understanding of AI risk is evolving as our use of technology. In that sense the biggest risk of AI is misunderstanding its potential risk. This can be caused by risks that occurred during operations that were not anticipated at design and are considered low probability but turn out to have catastrophic impact. One would need to consider elements such as autonomy, scope of ability to manipulate another system, novel context for use, novel data and degree of innovation.

Nevertheless, of all the potential risks we believe the highest risk evolves around autonomous or semi-autonomous AI systems that can have a physical impact either because of their sensory or physical properties. The risk relates to the impact such a system can have to cause physical damage or injury as well as the risk that the AI decisions will be inaccurate, unethical or open to manipulation by an attacker. The autonomous or semi-autonomous decision-making nature of the AI triggers the obvious liability questions and linked with that the issue of AI accuracy. We should remember however that quality of decision making in real life is not just determined by taking the right actions but also by omissions when a decision is made not to act. A system in Europe that is too restrictive and therefore trains AI inadequately may result into an AI that is failing to reach the right level of accuracy through its inaction.

How can one assess AI accuracy?

The determination of AI accuracy can happen on the basis of set parameters; However, to do it at one point in time only assumes a static (i.e. not evolving) AI. The moment the AI is evolving through self-learning the evolution of the results as well is depending on that learning. Part of the challenge is on the basis of whose data is the learning taking place? From the providers and manufacturer's perspective the AI system has been trained within certain parameters that guarantee a level of accuracy and explainability. The moment the AI system is put before customer data and is learning/trained using the customer data the properties of accuracy as a function of quality as well as explainability change to a point that the provider/manufacturer may also be unable to determine.

How likely is it to see customized AI and therefore uniquely trained?

Very likely. The current direction of the software industry is to make available AI within SaaS multitenant applications as a component of the overall solution that is there to serve each particular customer or a group of customers uniquely. Consequently, an AI tool begins its operations at given parameters of accuracy and explainability. The moment it comes in contact with real-life customer data and systematically processes them it will learn and evolve over a period of time. In the process It may acquire unique industry knowledge that are not visible to the manufacturer/producer or even to the customer and would turn out to be accurate and useful. For instance, an IT infrastructure management AI may turn out to have identified unique exposures on the deployment of Cyrillic alphabet keyboards because they could be associated with a particular vulnerability in an operating system. That "bias" would not be visible or understandable to the service provider/manufacturer of the AI. It would be something the customer's data is responsible for. It will probably not be visible to the customer either. In fact, it may even turn out to be useful in detecting more accurately errors in those Cyrillic language environments because the AI would statistically anticipate them. Such examples are likely to be the challenge in most AI implementations of real-life IT data sets, especially in use-cases of protective or management functions.

Data accuracy and AI

A lot of the discussion around GDPR-like principles highlights the importance of data accuracy and data quality as an indicator of quality for AI. Whereas in the context of personal data accuracy is an easier defined property in other use cases it may be more difficult. For example, in cybersecurity or fraud prevention the attacker by definition will supply inaccurate data. Therefore, a key capability of the AI needs to be managing inaccuracy. Producing accurate results is another requirement but as previously explained the accuracy will evolve over time and will change with the data supplied. Accurate data does not necessarily mean unbiased as real life often has hidden biases. The key requirement would be depending on the use case to factor certain biases that can result in unethical results and try to eliminate them. That would require however a level of governance and responsibility from the organization that is using the AI and supplying the real-life data on which the AI has trained and further developed beyond the capabilities of the manufacturer/service provider.

The criticality of accessing real life data

Any EU initiative in the space of AI needs to ensure that access to real-life data for AI training purposes remains unchanged and in fact more organizations have the right incentives to provide access to such data that would enable the training and development of AI. Supervised learning methods, which are critical in certain sectors such as financial fraud, cybersecurity, clinical diagnosis or infrastructure management depend on the availability of real-life data. The ability to develop effective AI in Europe that is qualitative and can securely be put into the market will depend on continuing to have access to such real-life information.

What about AI that is not processing data but develops applications or AI itself

One of the possible use cases for AI is the development of software applications or the development of new AI algorithms. For such use cases the AI training model is supervised with controlled data sets. Therefore, understanding the data quality that is used to train the AI and the explainability of its' decisions is an easier process.

What would be the impact of a certification/labelling framework?

Certification or labelling can be a cost because both require an upfront investment in effort and time to access a market. In fact, several compare certification/labelling with a market access tax. The ability of certification/labelling to generate a positive impact depends on conditions such as the cost and speed on certifying, the requirements to be met and their market relevance. In addition, one would need to balance the temporary assurance certification is providing with the overall slowing down of the product lifecycle and the diminishing assurance over time in cases of evolving AI. For example, the more certification/labelling delays the release of an AI capability on fraud detection, the more likely it is that by the time the AI reaches the market the modus operandi of the fraudsters to have changed. Equally, once a vulnerability in the AI is discovered, the more onerous the certification/labelling the less likely it would be for organizations to fix the vulnerability in fear of needing to go over a lengthy re-certification/re-labelling process before being able to use the AI tool, while the development of a patch would also be delayed in order go through a testing process that satisfies those certification/labelling requirements.

Quantification of the certification impact is based upon the investment (time to market and expertise) necessary for companies to conduct certification, versus the actual take-up of certified products by the market. It should be noted that any material product changes or versioning that needs to happen in order to meet the certification requirements in a particular region can increase the cost and therefore the negative impact of certification. In addition, certain certification requirements may result in some companies refusing to certify all together or negating a particular market. In a scenario that certification is mandatory it then becomes purely a discussion of cost and in particular whether the costs to certify outweigh the size of the market that the certification serves. If the market is large enough to justify the investment and certification is mandatory, then the certification cost is passed down to the market. If the market is large enough and labelling is voluntary, then the existence of labelling could function as a competitive advantage if it is widely recognized and adopted. In a discussion around the economics of certification/labelling the key question is that of market scale. The larger the market the certification serves the more likely it is for a positive investment case. Equally the least product changes/versions certification requires the lower the cost of producing certifiable solutions. This is especially important for AI technology that would be considered as state of the art.

The EU is trying to achieve a first mover advantage in this space. Therefore, certification/labelling requirements are likely to be more complex and the overall cost of certification to increase. This is why certification/labelling only for a limited number of EU Member States makes little sense in the EU context. The ideal scenario would be an EU labelling that has large scale adoption both in and outside the EU. It should be also noted that in order to achieve adoption both within and outside the EU the framework cannot be operating in a vacuum, but rather needs to take on board existing international best practices from other major players in this space.

Another risk of certification is the “freezing in time” effect. A certified system has a certain level of assurance, but it also cannot evolve. In the case of AI this is a major handicap as effective use of AI relies on data training and model evolution. In the case of other examples in the industry, such as medical devices, it has resulted in serious cybersecurity vulnerabilities in the ecosystem in an effort to minimize updating and the need for re-certification. Obviously such concerns apply also for AI.

Finally, on certification/labelling it is important to remember that establishing a certification/labelling framework for a technology that has so many different use cases and is so complex and evolving is very difficult. Therefore, making market access conditional on certification may result into a de facto market barrier for AI in Europe. Certification and standardisation are consultative and iterative in nature. The chances of seeing any meaningful wide-spread certification emerge before the technology gets into extensive real-life use are very slim, so setting that as a policy objective, and betting the success of AI rollout on that would be risky. Article 40 of the GDPR has been in force for 2 years, but even in spite of hard legislation existing, not a single European certification framework has been put in place to date.

One of the ideas considered as part of the Cybersecurity Act is that of a technical information scheme for basic assurance applications that would empower buyers and consumers to determine the security properties of the device/product they would acquire. The scheme could be based on self-declaration and would certainly be worth exploring as the regulatory mechanism to put it in place is already available.

Where do you believe the liability boundaries should be?

The discussion around AI and liability is a very complex one that should be studied in detail. A lot of the AI that is developed by manufacturers and service providers relies on algorithms and technology that they have sourced from others. However, as AI is not static and it evolves differently depending on its training the responsibility needs to be with different players in different phases of the AI operation and development. Therefore, liability and governance obligations should follow not just traditional models related with putting products into market or offering a service, but also follow the evolution and customization of AI as organizations deploy it and train it with data they control for their individualized use cases. Another point to consider is when reflecting on the liability boundaries is that there are scenarios where no ethical solution exists. For instance, the well-known scenario of self-driving car that needs to choose between the aging individual and the toddler is a scenario that has no acceptable ethical outcome.

What is the biggest risk if Europe gets AI policy wrong?

If, either because of excessive caution or overregulation, AI is not allowed to be (1) trained properly on real life data, (2) deployed at the appropriate time and (3) brought up to the relevant scale, then several major risks could materialize. Europe could miss out on the economic opportunity (which tries to capitalize through a first mover's advantage) by not being able to tap into the full potential of what AI could enable. This would put Europe at a strategic and structural disadvantage on the longer term. For instance, it could expose Europe to serious security and safety risks. For example only AI (as opposed to human analysts) can possibly manage efficiently the authentication and the security of automated digital transactions in a full scale 5G powered smart city. 5G is designed to enable connecting 1 million unique devices per square kilometer, which gives an idea of the number of transactions that would need to be executed reliably at any given time or over any given period in an area like Frankfurt, Amsterdam or Paris, from road and rail traffic control to energy and water distribution to sewage treatment and health infrastructure load balancing. This is without counting the billions of consumer gadgets and entertainment functionalities which, while not critical infrastructure, are a massive potential attack surface for malicious activity.

++++++++++++++++++++

We would like to thank the European Commission for giving us the possibility to provide feedback to this important issue that we believe will have a major impact in the development of technology policy in the European Union.

We remain at your disposal to provide additional information. Please feel free to contact:

Ilias Chantzos, LLM, MBA

Global Privacy Officer and Head of EMEA Government Affairs

Ilias.chantzos@broadcom.com