Guidelines on Ethics in Artificial Intelligence

American Chamber of Commerce to the European Union
Speaking for American business in Europe

Avenue des Arts/Kunstlaan 53, 1000 Brussels, Belgium • **T** +32 2 513 68 92
info@amchameu.eu • amchameu.eu • European Transparency Register: 5265780509-97

Introduction

The High-Level Expert Group on Artificial Intelligence

We welcome that the European Commission's High-Level Expert Group (HLEG) on artificial intelligence (AI) addresses important topics such as ethics and policy through an inclusive and multi-disciplinary approach. The objective of the HLEG should be to issue guidelines that are actionable and proportionate, which encourage companies to adopt a responsible and ethical approach to AI.

Around the world stakeholder organisations are developing high level ethical considerations that appear to be similar, but their weighting and actual implementation will vary according to geographies, industry and government sectors as well as business models. The more the HLEG's guidelines are tailored by these considerations through the inclusion of different use cases, the more its output will be relevant. Such tailored guidelines will allow companies to quickly start taking them into consideration in their research, development and deployment of AI systems.

The need for an ethical framework

There is no doubt that AI will transform every aspect of our society and economy. Already, AI is commonly used in many of our daily actions with features that make our live safer, more convenient and productive. Breakthroughs in areas such as vision, speech, translation and knowledge in recent years have proven AI is a technology that is transformative.

Because of this pervasive and transformative nature, it is sensible to also look at possible concerns about how the technology is developed (e.g. what human values, possible bias) and how it is deployed (e.g. what are the outcomes, purpose or application of the AI). Our objective should be to deploy and use ethics-embedded AI. For instance, there is a risk that AI systems perpetuate existing bias in society, willingly or not. We also need to be sensitive about the possibility that there are hidden assumptions and biases in data, and therefore in the systems built from that data — in addition to a wide range of other system choices which can be impacted by biases, assumptions and limits. This in turn can lead to outputs that replicate those biases and have serious blind spots. Business, government and other relevant stakeholders should seek to develop methods and leverage technology and other tools to detect and correct those errors and biases, and avoid replicating them.

It is nevertheless important to underline that ethical issues raised by new technologies like AI are also linked to its application and not only to the technology itself and its development. It is impossible for any technology that is operated by humans to remove all bias or be always entirely ethical. Furthermore, we should be careful in seeing AI as a potential solution to ethical issues arising from the deployment and use of technology in today's societies. Ethical goals and standard should therefore be reasonable and realistic, helping industry to strive while avoiding bias and ensuring an ethical application of AI.

We best address possible concerns by guiding the development and deployment of AI with ethical principles and a holistic AI framework which can include high-level principles, best practices, voluntary market and industry-driven standards as well as existing regulation. In the section below, we lay out what basic principles should be considered. These basic principles can be applied at several levels, starting with AI designers and developers. This could then be supplemented by tools such as a 'fact sheet' that provides basic information to AI users, depending on the AI application and sector.

Importantly, AI is no longer the sole domain of tech researchers but extends to users, philosophers, psychologists, sociologists and economists. This why a multi-disciplinary approach is needed to develop these ethical guidelines.

Considerations for an ethical framework

Addressing bias

AI systems should be designed in a way that the predictions resulting from training data are fair and as unbiased as possible. Because AI systems are designed by human beings and are trained using data that reflects the imperfect world in which we live, it is important that developers are aware about how bias can be introduced into AI systems and how it can affect AI-based recommendations. This requires responsible data governance and documentation, as well as regular testing to detect bias, including during deployment. It will be essential to train people to understand the meaning and implications of AI results to supplement their decision-making with sound human judgment.

Privacy & Security

Privacy is a key enabler of trust in all AI-related initiatives. As for all other technologies, AI systems must comply with privacy laws. This requires transparency about the collection, use and storage of data, as well as providing data subjects with appropriate controls.

Robustness

AI-based systems should operate reliably, safely and consistently. Demonstrating that systems are designed to operate within a clear set of parameters under expected performance conditions and that there is a way of verifying that they are behaving as intended under actual operating conditions is an important prerequisite for trusted reliability of AI systems. Design and rigorous testing should anticipate and protect against the potential for unintended system interactions or bad actors to influence operations, such as through cyberattacks.

Explainability, Transparency, Accountability

The starting point should be that users should know 'if' an AI system is being used, and 'how' and 'why' an AI system suggested a certain outcome over another. If we are aiming for trustworthy AI systems, transparency alone is only a first step. Ethical guidelines should also encourage the explainability of AI systems, with algorithms that are easily identifiable – from how they interpreted the task, to how they reached the output.

Transparency and explainability serve different purposes for different audiences at different times and are based on different machine learning techniques. There is no possible one-size-fits-to-all approach, therefore the notion must be tailored to different audiences, understanding for example that AI applications in business-to-business (B2B) and business-to-consumer (B2C) contexts imply different needs and levels of explainability required. Also, explainability will depend on the state of the technology, and what is technically possible at a certain time. If It is paramount that the outputs of the algorithms can be properly understood by non-technical audiences, explainability or intelligibility does not mean disclosure of algorithms or any intellectual property or trade secrets, which would do little to serve the needs of users.

By transparency we mean that, taking into account the technical possibilities, users should be provided with relevant, clear and meaningful information on what an AI system is intended to do, when it is being used, and

how outcomes are generated. Industry should develop meaningful ways to provide clear information on the composition and quality of initial datasets used to train original models. Depending on the applications, different levels of transparency and accountability might be possible. The ethical guidelines should encourage companies to develop these internal practices to embed certain ethical considerations in the development of models and in their application to products and services.

Considering the importance of accountability towards AI users, internal review boards could provide oversight and guidance on which practices should be adopted to help address the concerns discussed above and on particularly important questions regarding development and deployment of AI systems.

Success factors

In order to achieve their objectives, the general principles could be implemented in practice through:

- A framework that is proportional, risk-based and flexible because AI is not a monolithic technology but its ethical risk changes drastically according to its use and context.

- A 'holistic' framework including high-level principles, best practices, voluntary and industry-driven standards and existing regulation.

- A common understanding of what the problems are and further research to enable these problems to be more effectively addressed (e.g., technology can also be part of the solution).

- Encouraging companies to self-regulate: companies should establish guiding ethical principles for themselves that will apply throughout all their operations.

- Encouraging companies to adopt concrete governance practices: AI ethics should be built into business performance, not bolted-on as an afterthought. AI ethics should be part of the AI lifecycle, from the data models and product deployment, to the update of workflows, tools, and business processes. For example, companies could set up internal structures such as 'AI ethics boards' to discuss these issues.

- Encouraging companies to understand the key issues and tools to mitigate risk: As businesses and industry continue to pilot, adopt and rely on AI technologies to reshape the future of decision making, AI that can be trusted to be transparent, fair, explainable and secure is imperative. Businesses need to continuously listen to concerns that might exist and adapt their ethical guidelines in developing tools to mitigate risks.

- Continuous dialogue with stakeholders (industry, researchers, etc) on the development of appropriate mechanisms, in particular for any consideration of 'regulatory' mechanisms.

- Increase overall awareness and foster trust through the entire value chain, from developers to users, as well as consumers and society at large.

- Encouraging authorities to collaborate with industry and civil society in building data ecosystems which help to generate datasets in quantity and quality which ensure and empower a fair and ethical AI.

- Encouraging policy-makers to cooperate at international level on ethical guidelines, helping to ensure an inclusive and global approach.