Reply to Consultation

Commission White Paper

Artificial Intelligence - A European Approach

Prof. dr. Klaus Heine & Prof dr. Evert F. Stamhuis LLM

Erasmus School of Law, Erasmus University Rotterdam

Co-director and senior fellow Jean Monnet Centre of Excellence on Digital Governance

www.digov.eu

11.06.2020

Introduction

The White Paper addresses a number of highly important issues regarding the upcoming digital technology, that uses AI in its various forms of appearance. The Commission and Council would be right in taking its own strong position in the global discourse. In that way Commission and Council give strategic guidance for the European economy and at the same time display to European citizens that the Union stands up to protect the values that represent our common European perceptions of a good life.

In this reply we limit ourselves to comments on three issues, in which we partly endorse the approach the Commission advocates and give guidance regarding the follow-up. *Firstly*, we address and support the sectoral approach and comment on the proposal for a governance/enforcement structure. *Secondly*, we proceed to deal with liability issues, where we point at the challenges to overcome in the era of New Technologies. *Thirdly*, we put the human centeredness of AI systems as building block for the ecosystem of trust to the test, thus demonstrating how a contextual approach leads to the best possible results.

Sectoral versus generic governance

The risk-based approach leads the Commission to propose a sectoral governance, with priority for high-risk sectors. We raise the question whether the presupposed risk assessments properly identify the high-risk sectors. It is not sufficiently clear what the concept of risk comprises and what not. Risk is a multifaceted concept in which chances for failure, chances for damage, nature/seriousness of damage and chances for loss of trust can all be components. Furthermore, even when the concept is clear, the appreciation of the chances is highly influenced by political views and economic interests. So, we have doubts whether the risk variation is a widely acceptable guidance to lead us to the high-risk sectors. Moreover, inside a sector a so-called innocent AI application may maintain or exacerbate social inequalities. At least the Commission should come forward with additional evidence regarding the risk assessment that will guide the prioritizing process. Given the yet not clear meaning of risk (and its different perceptions) regarding AI it is recommended to employ risk categories that allow for future adaptation.

This does not prevent us from endorsing a sectoral approach to governance of AI. In our view, the commonality of the used technology is insufficient to support a generic approach. In no way does the AI-governance topic resemble the issues that the EU has generically dealt with, such as data-protection. The GDPR could be produced as a technology-insensitive piece of regulation for the protection of ingrained European values, importantly building on decades of practices developed under the preceding governance regime. But AI does not represent a common European value, it is a technological novelty. It is very impactful and appears in several social and economic sectors of our society. Those sectors have in the past managed with the upcoming of other disruptive technologies. The sectors changed their ways of working and communicating and adapted to the new conditions. We are not convinced that AI is in that respect *fundamentally* different.

AI will have great impact on many aspects of our lives. But it is for society to determine when and where that happens and to what extent. We resist any form of technological determinism. The social impact of AI is not an inevitability, it is a policy decision on the level of individuals, families, companies/employers, doctors and nurses, public administrators, governments, etcetera. Hence, the governance should be tailored towards the concrete effects on the interests that are impacted. That can only be properly assessed, discussed and decided at the lowest level possible, where the consequences of policy choices play out in real life. It is obvious that general high-level principles give some guidance (such as the HLEG principles for ethical AI), but it is also undoubtable that in many instances the principles give contrarious guidance. That implied controversy can only be resolved by the persons or institutions that will experience the consequences of the choice for solution A or solution B. This implies that in our view that the governance should reflect to the extent possible the specificities of the interests at stake, which cannot be expected to happen in a meaningful way across sectors. It is not the technology that dominates the governance issues, it is the concrete mix of interests that must be balanced in the governance approach. This leads us to underscore that a sectoral approach is an inevitability.

In this respect we find the White Paper lacking in supporting the necessity of a structure at the Union level as instrument of governance. It appears that a structure familiar to the one for data-protection

is aimed at. A single institution at the Union level combined with national agencies could indeed provide for the necessary generic expertise and weigh in seriously against other actors when executing its duties, such as professional associations, tech providing industry etc. But we do not see convincing evidence that this institutional framework, like the one for data protection, is the panacea to overcome all governance diversification. It is a continuous pulling and pushing to achieve sustainable coordination and cooperation with sectoral institutions, at least at the member state level. Moreover, the trade-off would be a greater distance towards the specific sectors, resulting in a less effective governance culture. Therefore, we encourage the Commission to investigate alternative routes for coordination, communication and joint expertise, so that in a possible sectoral structure contradictory and mutually damaging policies are prevented.

Regulatory *agencies* can be established tailor-made for a specific sector and purpose and would be staffed with experts from the field. For example, an agency could analyse algorithms for artificial intelligence systems in health care; like the FDA does in the USA. That way EU-level agencies concerned with New Technologies would strongly support the level playing field in the *Single Market*.

One important aspect with regard to those sectoral agencies is their relationship with the legislator. A legislator typically fears losing control over agencies ("mission creep") and therefore it tends to restrict the flexibility of agencies. In case of New Technologies this leads to a trade-off that has to be wisely balanced: On the one hand any agency regulating New Technologies needs a strong legitimacy back-up by the legislator, on the other hand there must be room for flexibility of regulation, because of the technological dynamics and to prevent a situation in which innovation and growth become stifled.

Liability issues

The aim of any liability regime is twofold: 1) In the event of an accident the victim shall get a fair compensation. 2) The liability regime shall allocate responsibilities along the value chain from the producer towards the consumer. The producers and consumers shall be incentivized to take due care, thereby implementing the optimal care level. Under ideal circumstances a liability regime leads to *optimal deterrence* of producing or selling defective products.

While liability regimes around the world may differ in detail, it is safe to say that they have always adapted to the technical givens and societal circumstances relevant at a specific moment in time. Over the last hundred years we see as a general pattern that mass manufacturing led first to the legal innovation of *organizational liability* and later to *mass torts* in which whole industries became liable (for example, asbestos cases). Moreover, the EU accomplished in the Product Liability Directive not only a high standard consumer protection, but the Directive also helped to fulfil the Single Market by creating a level playing field amongst European companies.

In the past the developments in liability law were challenged mainly from a doctrinal legal point of view. Thereby the legal developments built upon each other and could rely on an analytical framework in which it was clear who were the producers and who the consumers. There was also certainty about the properties of the production technology, the single steps in the value chain and

whom to hold responsible in case of wrongdoing. However, this is no longer the case in a world of big data, artificial intelligence and 3D-printers. The advent of New Technologies leads to a couple of *legal disruptions* which make it necessary for the EU to reconsider the incumbent liability regime and to replace (if necessary) incumbent laws, regulations and procedures.

The producer-consumer divide

In incumbent liability regimes around the world it is (rightly) assumed that there is a clear-cut distinction between producers and sellers on the one hand and consumers on the other hand. Liability regimes typically target producers because they are the ones who profit mostly from business, who can pay out compensations and who – most importantly – are the ones who can make by the help of research and development the product safer. Moreover, producers have typically a (large) production site which makes it possible to physically spot the issuer of a defective product.

In the world of New Technologies, the distinction between producers and consumers has faded away with the emergence of the so-called *prosumer* as new category. This makes it far more complicated to target the manufacturer of a defective product, and which makes the incumbent tort law a blunt instrument.

For example, a hobbyist software-programmer may engage in co-drafting an open-source code for an artificial intelligent entity on a platform with others. Afterwards the hobbyist programmer uses the entity by herself, and others use the source code for commercial purposes. If the artificial intelligent entity creates accidents, is it then the hobbyist programmer who can be held liable?

In such cases it is questionable whether hobbyist programmers are strictly liable according to product liability, because they are also consumers and have no essential business or financial interest in the activity. Nevertheless, the activity may unfold quite a large impact on business activities (for example, an open source code may become quite popular for any kind of programming). But even if one would believe the hobbyist programmers or producers can be made strictly liable within the incumbent liability law, the two primary goals of liability law will not be reached. 1) Typically, the hobbyists do not have the funds to pay compensations as companies can do. 2) The deterrence of the liability regime will and cannot incentivize the hobbyists to engage in systematic research and development, in order to improve product safety as it is the case with companies. That means the two main reasons for the incumbent liability law do not longer work, because the divide between consumer and producer is fading away in the realm of digitization.

The question is then into what direction the liability regime could be further developed to effectively protect consumers on the one hand and to facilitate New Technologies on the other hand.

A first measure is to complement incumbent product liability law by specific *design regulations*. That means to prescribe or to forbid certain technical properties that may create hazardous or defective products. However, the probability of product failures depends in most cases on the digital source code (and algorithms) and cannot be controlled by design regulations of the hardware alone. Moreover, regulators have neither the capacity nor the capabilities to assess source codes regarding potential defects.

Given that background, and in line with the two aims of liability regimes, it would be straightforward to hold the *digital platforms* (or intermediaries) liable at which the exchange of source codes and/or digital designs takes place. There is a clear advantage of this strategy, because platforms have or can build up the capacity to assess the potential dangers of source codes. Ultimately, they could deny the access to the platform. A platform may also have the financial capacity to pay compensations. Moreover, it is much easier to locate a digital platform on the internet than a single contributor to a platform. In short, digital platforms may take over the responsibilities that actual producers have according to incumbent liability regimes. Along this line of reasoning consumer protection could be further strengthened if platforms would need a license for doing business. Here the EU level could play an important role to guarantee high and EU-wide (legal and ethical) standards. Moreover, such a licensing at the European level could preserve a level playing field in the *Single Market* stimulating competition. This might be a decisive step for fulfilling the *digital single market* enshrined into an analytically coherent *digital governance* at the EU level.

Finally, a third measure is to introduce compulsory liability insurance for specific activities at the household level where AI systems play decisive roles in daily life. In the European Union this is yet already the case with regard to the operation of drones, to which we refer for inspiration for concrete legislation.

No soul to damn, no body to kick

“Did you ever expect a corporation to have a conscience, when it has no soul to be damned, and no body to be kicked?" (Baron Edward Thurlow, 1731-1806). This famous quote was once targeted at corporations when fitted with their own legal personality. It was assumed that corporations - other than humans - would create harm to society by financial fraud and undertaking hazardous activities. And indeed, organizational (enterprise) liability became a challenging topic since the times of industrial mass production.

With regard to AI the issue of *legal personality* enters again forcefully the agenda. This is for two reasons. 1) If an artificial intelligent system takes decisions independently from a human and the human has moreover no insight into the artificial decision-making process, then it is no longer possible to assume a human as responsible for the decision. The White Paper already referred to this complication. 2) An artificial intelligence cannot be deterred; it has neither a conscience nor can it be sent to jail or would feel a loss of utility by being fined.

Maintaining human responsibility for an artificial intelligence through ownership is in principle possible but its effect remains very limited in terms of the two goals of liability law. Associating an AI with a human through ownership may solve the problem of compensation if the owner is a company. But ownership does not solve the problem of a lack of deterrence, because the AI would still act independently from the human who has no insight into the decision-making process of the AI. To make this even more poignant: criminal law is not applicable to an AI and cannot prevent it from taking wrong action.

Therefore, it might be more appropriate to accept that an AI may have a sort of legal personality, accepting that humans cannot be put in charge of the decisions of an AI. This does not mean that AIs

are humanlike, but that certain rights and obligations can be directly attributed to AIs. Instead of an all-encompassing *property right* a *contractual* relation would constitute the association between humans and AIs.

Granting legal personality to AIs may lead the discourse about the legal status of AIs into a new direction that does not only hint to smart solutions for liability issues but connects also to other legal areas, particularly *intellectual property* and *competition law*. For example, it can be assumed that artificial intelligence will create new technological solutions, like algorithms, pharmaceuticals, integrated circuits, energy supplies etc. If the intellectual property rights are (partly) with the AIs themselves society has a better chance to fully exploit the value of those inventions and to facilitate fair competition among those products. Otherwise, society must deal with hardwired property rights that vest with the owner of the AI and which might be abused to exert market power. Hence, the question about the legal personality of artificial intelligence has also repercussions into competition law.

In sum: we plead for a reconsideration of the proposals that among others the European Parliament has adopted, taking into consideration the opposition it engendered.

Disrupted value chains

New Technologies are constituent for the Internet of Things, Smart Factory or Industry 4.0. The Internet of Things makes it possible that in one country a smart algorithm produces a CAD-file which is then directly sent to a 3D-printer in another country, where the printed product instantly can be sold. This example contains a couple of legal challenges. With regard to liability regimes two issues play the foremost role: *Law enforcement* and *generic security*.

Again, making platforms (intermediaries) liable for the surveillance of potentially harming products is an option. Another possibility might be the registration of natural persons as owners of specific hardware (for example, 3D-printers or even data centres). China is already implementing such a strategy to better trace back responsibilities. However, international law enforcement in the world of Internet of Things will remain a major challenge for the future.

Regarding generic security the disruption of value chains creates further challenges for society. The regular control mechanisms for access to a local market do not provide any longer. In the case of 3D printing the digital files in combination with 3D-printers do no longer require the production of intermediate goods which can be physically controlled along the value chain. As with 3D printing, the final AI application is an accumulation of components where violations of the local laws and standards can remain hidden in the process. For example, facial recognition software, deployed by a recruiter in the EU, can be produced in India, where the algorithm was trained on data, that were collected from scrapping social media against our legal and ethical standards. The recruiter will comply to GDPR standards for the data on which he releases the model, but the model itself has an illegal production history that will stay hidden in the chain.

Beyond the liability issues that unfold because of disrupted value chains in the wake of New Technologies there are two further challenges which have entered the agenda and are related to liability issues. 1) If there are less possibilities for border controls then this implies that there are also

less (intermediate) products that are shipped. As a consequence, harbours will possibly play less a role for transhipment of (intermediate) products. At the same time the shipping of raw materials will become more important. 2) Governments are making revenue along value chains by putting Value Added Tax and tariffs on goods, correspondingly to the added value at a step in the value chain. In the Internet of Things it is no longer clear when and where the taxable added value has been created. Is it the CAD-file or algorithm created in one country or is it the 3D-printout or service delivered in another country that creates the tax base? In any case the disruption of value chains will make it necessary to come up with new revenue sharing rules between countries. The EU should act as a motor for such rules that support fair revenue allocation between jurisdictions.

Ecosystem of trust with human-centered AI

The White Paper promotes an ecosystem of trust, for which an important building block is to place the humans center stage. This is indeed in line with the ethical guidelines that the HLEG has shared. However, we must admit that this raises as many new questions as it answers. We believe that the answer to the question "What does it mean to say, that humans are at the center?" can only meaningfully been given, if one approaches the matter at the concrete level. That is where the AI system has its impact. So, let us practice that regarding using human-centered AI technology in public administration; one of the *high-risk sectors* in the view of the Commission. The question is then rephrased as: "What does human centeredness bring the individual as entitlements towards the pubic authority, whenever she is confronted with an AI powered system?"

It is far too simple to say: she always has the right to ask a personal intervention by a human. Effectively that right only exists in special circumstances (such as under art. 22 GDPR). Moreover, what is the sensibility of that entitlement, when a human intervention increases the chances for biased decisions, negligence and delays? We have not yet seen evidence of real and meaningful benefits in the human intervention as correction mechanism for automated processes. The theoretical hypotheses lead us to a mixed expectation. In the psyche of that intervening human the tendency to execute autonomy will struggle with the tendency to agree to the system, which tendency is reinforced every time the machine turns out to have given a correct outcome. On top of that, organizational efficiency may press the human intervenor to correct as little as possible. So, until now we do not maintain high expectations for the intervention to go beyond rubberstamping the machine outcome. One of the proposed assets of AI is that a personalized but neutral approach is better safeguarded than when cases are in the hands of individuals who take repetitive decisions. Hence, the entitlement should not be absorbed by allowing/prescribing human intervention.

In addition, one cannot say that the individual has an entitlement to stay undiscovered, particularly when authorities use technology to improve proper use of public funds; or combat a serious health crisis like the current COVID-19 pandemic. There is no such thing as a right to violate the law and stay under the radar. There is indeed a right to challenge the evidence and the sanction, whenever the individual is submitted to such a response. Also, the right not to be subjected to discriminatory scrutiny is recognized. And there is the right to be forgotten and erased from the memory of the machine, at an appropriate time.

In our view, the individual is entitled to be treated as individual, to the extent necessary for the specific context, where human dignity requires respect for the capacities as well as the fallibilities. This is true for treatment by an AI powered machine to the same level as it is for treatment by a human. This brings some specific requirements for AI implementation: 1) The design and implementation should attain a maximum level of protection against unlawful discrimination and biases, where unlawfulness is determined by concrete policy contexts. For example, a system of social benefits has the basic intention to correct inequalities in economic circumstances, where a subsidy scheme for employing the lesser able has the purpose of adjusting the labor market. In different contexts the consideration of different inequalities will be lawful. 2) The deployment and operation of the AI application should be made transparent up to a level that allows for an effective challenge of the outcome if necessary. This is a direct consequence of safeguarding access to justice, so that respect for human rights can be put to the test in concrete cases. This is to a large extent absorbed in the elaboration of the requirement of explainable AI. 3) The system should not close the door to the application of hardship clauses in individual cases. In that respect AI implementation may have the potential to turn back the clock somewhat for individual citizens. Hardship clauses should be there to mitigate those potential effects of unduly strict rule-application and enforcement; and in that way preserve a human face in public government. Moreover, these clauses give way to the presence of excusable failure on the side of the individual. Not every person is sufficiently literate, let alone digitally literate. Estimations reach levels of 3 % of the population with regard to functional disability in reading and typing. To be clear: this plea for effective space for hardship does not necessarily rule out automation. The best possible provision to apply hardship clauses is defended here, be that a human or a sophisticated AI powered automat.

Conclusion

This concludes our reply to the White Paper for now. We partly supported and partly criticized the content of this document. In addition, we pointed at issues for further policy development. As the Jean Monnet Centre of Excellence on Digital Governance we remain available, should our reply require further explanation, or when any other assistance is sought.