FEEDBACK ON THE INCEPTION IMPACT ASSESSMENT

Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence

September 10, 2020

Virginia Dignum
Catelijne Muller
Andreas Theodorou

ALLAI.

In its Whitepaper on Artificial Intelligence, Europe took a clear stance on AI; foster uptake of AI technologies, underpinned by what it calls 'an ecosystem of excellence', while also ensuring their compliance with to European ethical norms, legal requirements and social values, 'an ecosystem of trust'. The Inception Impact Assessment of a "Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence" now presents a number of objectives and policy options. This paper provides feedback to these objectives and policy options.

By Virginia Dignum, Catelijne Muller and Andreas Theodorou

Executive Summary

First and foremost, we would like express our support for the European Commission's efforts to establish an appropriate regulatory framework for AI. In establishing such a framework, one should both look at existing laws and regulations and determine if their are 'fit for purpose' for a world with AI as well as consider establishing new rules where current legislation is not adequate.

In general, we recommend to broaden the description of the problem that the initiative aims to tackle, i.e. *addressing a number of ethical and legal issues raised by AI*, and include "*societal* issues raised by AI". In the same spirit, we recommend to broaden the description of the ultimate policy objective of the proposal, *i.e. to foster the development and uptake of safe and lawful AI that respects fundamental rights across the Single Market by both private and public actors while ensuring inclusive societal outcomes*, so as to include "*fair* societal outcomes".

The issue of defining the scope of a new legislative initiative for AI is the core element that needs to be addressed. Whereas the Inception Impact Assessment mentions a number of AI-techniques that either should or should not be covered by the instrument, we would like to recommend a different approach toward defining the scope of the instrument: an approach that looks at the level of impact of the technology on people and society at large, rather than (merely) on the technical specifications of a particular AI-system. An impact-level based approach lowers the risk of loopholes that could be exploited.

As for existing legislation, we call for a broad legal AI stress test, because we see a large number of additional legal lacunae where it comes to AI, that were not mentioned in the Inception Impact Assessment, such as the GDPR, law enforcement, competition law, transportation, trade of dual use technology, medical devices, energy and the environment, to name a few.

The Inception Impact Assessment lays down 5 policy options ranging from keeping the 'Baseline scenario' to a combination of several policy options.

ALLAI would be most in favor of a combination of the policy options 2, 3a and 3b as described in the Inception Impact Assessment. This combination would entail soft law for low impact AI applications (or uses) including voluIntary labelling, and EU instrument with mandatory labelling covering two elements: (i) clear restrictions, conditions, safeguards and/or boundaries for a limited number of exceptionally impactful AI-applications or uses and (ii) mandatory requirements for medium to high impact AI based on common denominators to determine the level of impact.

Finally, we call for an *ex-durante* (which would include *ex-ante* and *ex-post* mechanisms) mechanism to ensure a continuous, systematic, socio-technical governance approach, looking at the technology from all perspectives and through various lenses. For this we recommend to set up European AI Authority as part of a global framework of AI Authorities.

A. Context, Problem Definition and Subsidiarity Check

1. Defining AI for regulatory purposes

While we realize that the Inception Impact Assessment is not the place to provide for a definition of AI, it rightly identifies the need to define the scope of the initiative as a core question that needs to be answered. We would like to make some remarks on the issue of defining AI for regulatory purposes.

As we already indicated in our "Analysis of the EU Whitepaper on Artificial Intelligence"[1], we want to emphasize that AI is more than data and algorithms, powered by Computer Processing Power (CPU). While this is the case for the most widely used AI-systems at present, this is only a very limited description of what AI is. AI is a container term for many computer applications, some of which combine data and algorithms, but other, non-data-driven AI approaches, also exist, e.g. expert systems, knowledge reasoning and representation, reactive planning, argumentation and others. This should be kept in mind when defining the scope of the initiative, to avoid over- or underinclusion.

In the same analysis we have elaborated on the fact that there is no universally accepted definition of AI and how we would describe AI technically. Nevertheless, it should be noted that many of the wide applied AI systems are indeed examples of data-driven AI, that have a number of typical charachteristics that can make them brittle, unstable and unpredictable (such as opacity, correlation in stead of causality, insufficient or low quality data, unclear 'goals', a lack of common sense, etc.).

[1] https://allai.nl/wp-content/uploads/2020/06/ALLAI-Final-Analysis-of-the-EU-Whitepaper-on-AI-consultation.pdf

Legal versus technical definition - a need to look at the impact of AI

It is important to realize that legal definitions differ from purely technical definitions whereas legal definitions should meet a number of different or additional requirements[2] such as inclusiveness, preciseness, comprehensiveness, practicability, permanence, some of which are legally binding, and some are considered good regulatory practice[3].

In general, we feel obliged to emphasise that the focus on a definition through any attempts at defining AI-techniques in order to determine *what is* and *what is not* AI creates loopholes that could be exploited. This brings us to a number of important groundrules that should be guiding AI regulation and how one should set the scope of such regulation or, in other words, define AI for regulatory purposes:

1. We recommend to focus on the effects and impact of the systems, not on a particular AI-technology/technique.
2. AI-systems are more than just the sum of their technical or software components. AI systems also comprise the socio-technical system around it. When considering regulation, the focus should not just be on the technology, but more on the social structures around it: the organisations, people and institutions that create, develop, deploy, use, and control it, and the people that are affected by it, such as citizens in their relation to governments, consumers, workers or even entire society.
3. An "AI lifecycle approach" should be followed, that considers not only the development stage of AI, but also the deployment and use stages.
4. It should be kept in mind that most AI-applications currently being used could enshrine, exacerbate and amplify the impact on existing laws and fundamental rights as well as society at scale, affecting larger parts of society and more people at the same time.

ALLAI Project: Defining AI for Regulatory Purposes

ALLAI is currently evaluating the feasibility of setting up a project called "Defining AI for Regulatory Purposes". Working with a small group of experts and scientists from different backgrounds the project aims to provide guidance for the open core question relating to the scope of the initiative, notably how AI should be defined.

[2] A Legal Definition of AI Jonas Schuett Goethe University Frankfurt September 4, 2019 *(Legal definitions must be: (i) inclusive: the goals of regulation must not over- or under-include. (Julia Black. Rules and Regulators. Oxford University Press, 1997. [32] Robert Baldwin, Martin Cave, and Martin Lodge. Understanding Regulation: Theory, Strategy, and Practice. Oxford University Press, 2nd edition, 2012.); (ii) Precise: it should be clear which case falls under the definition and which does not; (iii) Comprehensive: the definition should be understandable by those who are regulated; (iv) Practicable: legal professionals should be able to easily determine whether a case falls under the definition; (v) Permanent: the need for continued legal updating should be avoided.*

[3] Inclusiveness can be derived from the principle of proportionality in EU law (art. 5(4) of the Treaty on European Union. The criteria precision and comprehensiveness are based on the principle of legal certainty in EU law. The criteria practicability and permanent are considered good legislative practice.

2. Problem the initiative aims to tackle

On: "Ultimate objective"

ALLAI for the main part supports the ultmate policy objective of the proposal, *i.e. to foster the development and uptake of safe and lawful AI that respects fundamental rights across the Single Market by both private and public actors while ensuring inclusive societal outcomes.*

ALLAI would however advise to broaden the final part of the ultimate policy objective to not only ensure *inclusive* societal outcomes, but to also ensure "*fair* societal outcomes". As mentioned in the Inception Impact Assessment, the complexity (and many times opacity) of certain systems and granular applicability of outcomes to individuals in combination with the scalability of AI systems, presents a range of difficulties as regards enforcement of existing legislation meant to protect human rights and could generate new safety risks. On top of that, these characteristics could enshrine, exacerbate and amplify these risks and adverse impacts on society at scale, affecting more people at the same time. As such there is a serious risk that unfair societal outcomes become ever more enshrined, exacerbated and amplified, thus potentially leading to wider and deeper societal gaps between groups of people, propagating inequality and, as a consequence, entrenching political polarisation.

The addition of fairness to ultimate policy objective would also reflect the European Commissions' adoption of the 7 Requirements of the High Level Expert Group on AI, i.e. requirement no. 5: Inclusiveness, non-discrimination and fairness.

On: "Harm caused by AI-systems and risks not covered by existing legislation"

We agree with the indication that harm caused by the use of AI may be the consequence of multiple causes. We would like to stress however that the causes for harm go beyond just flaws in the technical and digital components (including data) and charachteristics of the system. While these flaws do often play a major and sometimes decisive role in causing harm, one should not forget that even the most technically robust systems can still cause harm.

Imagine a facial recognition system that does recognize people of all colors, genders, ages, etc. correctly. Technically such a system could be considered robust and non-discriminatory. Lawfully however, the system could still cause harm. AI-driven (mass) surveillance with facial recognition, involves the capture, storage and processing of personal (biometric) data (our faces), but it also affects our 'general' privacy, identity and autonomy in such a way that it creates a situation where we are (constantly) being watched, followed and identified. As a psychological 'chilling' effect, people might feel inclined to adapt their behaviour to a certain norm, which shifts the balance of power between the state or private organisation using facial recognition and the individual.

In legal doctrine and precedent the chilling effect of surveillance can constitute a violation of the private space, which is necessary for personal development and democratic deliberation. Even if our faces are immediately deleted after capturing, the technology still intrudes our psychological integrity.

This is just one example where the mere technical elements of a system are less relevant as regards harm than the actual use of the system. That is why we advocate to continuously ask "question zero": Do we want to allow this particular AI-system and technique in the first place, or are there reasons not to allow its use at all? And if we were to consider implementing, deploying and using such a system, what are the conditions we should set for its use?

In either case, for the purposes of any regulatory framework we should not merely focus on technical solutions at dataset or algorithm level, but devise socio-technical processes that help us:

a) Understand the potential legal, ethical and social effects of the AI-system and improve our design and implementation choices based on that understanding;
b) Audit our algorithms and their output to make any undesirable outcomes transparent; and
c) Continuously monitor the workings of the systems to mitigate the ill effects of AI.

On: "Risks not adequately covered by existing legislation"

ALLAI agrees with the Inception Impact Assessment that there are ample risks that are not adequately covered by existing legislation on cybersecurity, protection of employees and anti-discrimination. The Inception Impact Assessment identifies as main issues:

- Effective enforcement of existing EU rules to protect fundamental rights
- Application of EU rules on safety
- Application of EU the rules on liability

AI & the impact on fundamental rights

ALLAI strongly commends the focus on the protection of fundamental rights and would like to draw your attention to a report it delivered to the Council of Europe on the "Impact of AI on Human Rights, Democracy and the Rule of Law".[4] This report identifies those human rights, as set out by the European Convention on Human Rights ("ECHR"), its Protocols and the European Social Charter ("ESC"), that are currently most impacted or likely to be impacted by AI. It aims to provide a number of possible strategies that could be implemented simultaneously, ranging from addressing the impact within the existing framework of human rights, democracy and the rule of law to establishing new human rights should the existing framework fail to adequately protect us.

[4] https://rm.coe.int/cahai-2020-06-fin-c-muller-the-impact-of-ai-on-human-rights-democracy-/16809ed6da

While states are obliged to protect individuals and groups against breaches of fundamental rights perpetrated by other actors, appreciation of non-state actors' influence on human rights has steadily grown.[5] As (large) tech companies have now become operators that are capable of determining and perhaps altering our social and even democratic structures, the impact of their AI(-use) on fundamental rights becomes more prevalent. In this respect, AI might serve as a good opportunity and think of a structure that goes beyond the 'horizontal effect' of the EU Charter of Fundamental Rights of the European Union. Such a structure could entail a legal obligation for private actors to comply with fundamental rights and to grant access to justice if they fail to do so.[6]

AI & the impact on work

The Inception Impact Assessment mentions that there are legal issues as regads the protection of employees, but does not yet specify these issues. We would like to stress that he introduction and use of AI in the workplace can cause effects as to health and safety in the workplace, job security, worker privacy, the balance between worker and employer and so on. For that reason, we have been advocating early and close involvement of workers and service providers of all types, including freelancers, the self-employed and gig workers – not just people who design or develop AI, but also those who purchase, implement, work with or are affected by AI systems. Social dialogue must take place before the introduction of AI technologies in the workplace, in line with the relevant applicable national rules and practices.

Additionally, we would like to draw special attention to AI used in hiring, firing and worker assessment and evaluation processes. The White Paper on AI mentions AI used in recruitment as an example of a high-risk application that would be subject to regulation irrespective of the sector. We recommend extending this use to include AI used in firing and in worker assessment and evaluation processes, but also to explore the common characteristics of AI applications that would make for a high risk use in the workplace, irrespective of the sector. AI applications that have no scientific basis, such as emotion detection through biometric recognition, should not be allowed in workplace environments.

Additional legal lacunae

In addition to the issues identified in the Inception Impact Assessment, ALLAI also identifies a number of additional regulatory lacunae related to existing legislation.

As AI is evolving quickly and the wider impact of AI on the full *acquis* has still not been fully identified, these lacunae are not exhaustive but below are the areas for attention to the extent that they can be identified today[7]:

[5] Business and Human Rights, A Handbook for Legal Practitioners, Claire Methven O'Brien, Council of Europe
[6] This also means going beyond merely referring to the Recommendation CM/Rec(2016)3 on human rights and business of the Committee of Ministers of the Council of Europe (and the UN Guiding Principles on Business and Human Rights)
[7] See also: STOA Policy Briefing: Legal and ethical reflections concerning robotics, The Scientific Technopolis Group, Maastricht University, Scientific Foresight Unit (DG EPRS) of the European Parliament.

1. GDPR
2. Law Enforcement
3. Competition law
4. Transportation
5. Trade of dual-use technology
6. Consumer protection
7. Healthcare
8. Energy & Environment

Ad 1: GDPR

A much cited existing EU regulation in the context of AI is the GDPR. This regulation provides frameworks for data protection, the right to an explanation of AI decisions and safeguards for the use of biometric recognition.

Biometric recognition: It should be noted that GDPR restricts the processing of biometric data only to some extent. Biometric data according to the GDPR is "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, *which allow or confirm the unique identification of that natural person.* The last part of the sentence is crucial, because if biometric recognition is not aimed at identification (but for example at categorization, profiling or affect recognition), it might not fall under the GDPR-definition. In fact, recital 51 of the GDPR says that 'the processing of photographs [is considered] biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person.'

Many biometric recognition technologies are however not aimed at processing biometric data to uniquely identify a person, but merely to assess a person's behaviour (for example in the classroom) or to categorize individuals (for example for the purpose of determining their insurance premium based on their statistical prevalence to health problems). These uses might not fall under the definition of biometric data (processing) in the GDPR.

Right to explanation of ADM, including profiling: Particularly with regard to the right to an explanation of automated decisions, a debate is underway about whether the GDPR gives a right to an explanation of automated decisions or not. Under the GDPR, controllers who use personal data to make automated decisions are obliged to inform individuals in advance and to provide meaningful information about the logic and importance of the decision-making and the consequences for the data subject. The so-called art. 29 Working Group recognizes that "the growth and complexity of machine learning can make it challenging to understand how automated decision-making or profiling works," but that, despite this, "the company [must] find simple ways to tell the individual about the reason, or the criteria on which the decision is based, without necessarily always attempting a complex explanation of the algorithms used or disclosure of the full algorithm."

Also, the GDPR requires a controller to implement appropriate safeguards when designing automated decisions, such as the right to human intervention and the right to express his or her point of view and contest the decision. The recitals of the GDPR also include the right to an explanation of a fully automated decision.

It is however debatable whether there is a full right to an explanation of an automated decision in all cases. The fact that the right to explanation is included only in the recitals seems to be a hurdle that can be overcome, but in particular the fact that the right to explanation only exists as regards to fully automated decision-making, makes it insufficient to ensure adequate transparency in the automated decision-making process. After all, a single human 'check' on an automatic decision (regardless of whether this person has been able to judge the decision on its merits, could lead to the conclusion that an explanation within the meaning of the GDPR would not be necessary.

It must also be assessed whether the GDPR offers sufficient protection when the decision is based on non-personal data. AI-driven profiling and the categorization of people or groups of people is regularly being done by finding inferences made about an individual, even without using personal data or resulting in identification of a person. There is no consensus whether these inferences in itself should count as personal data, but there are experts arguing that all data processing that has an impact on people should be protected[8] or even that the distinction between personal and non-personal data should be suspended[9].

Here the issue also ventures from the GDPR into other fundamental rights such as the right non-discrimination, covered in multiple EU directives and the ECHR.

Ad 2: Law enforcement

In line with the above, EU regulation such as the Police Directive (EU 2016/680), that regulates the processing of personal data in particular for the purpose of profiling by law enforcement in the Member States, does not adequately cover the issue of profiling that is based on mere inferences rather than personal data, as described above.

Ad 3: Competition

Many AI-applications are developed and deployed by a handful of private actors. These actors are also present in multiple market segments that are related to AI or use AI, such as finance, insurance, etc. If too much market power in these different markets is concentrated in a few companies, this could lead to unfair competition and difficulty for new (smaller) players to enter the market.

[8] Purtova (2018)
[9] Koops (2014)

Ad 4: Transportation

While autonomous driving has not yet fully materialized, the autonomous vehicles industry is well underway to making vehicles behave more autonomous and needing less 'active involvement' from their drivers.

These developments call for a critical review of the various EU regulations and directives that deal with transportation, such as Regulation (EC) 561/2006 and Regulation (EEC) 3821/85 regarding driving and resting times and digital tachygraphy (for truck platooning), Directive 2014/45/EU on Roadworthiness, Directive 2010/40/EU on Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport and Directive 2003/59/EC on training and initial qualifications of professional drivers[10].

Ad 5: Trade of dual-use technology

Following an impact assessment in 2016, a new reform process was started to (a.o.) future-proof the export control regime for rapidly developing emerging technologies. In sum, the main goal of this reform is to control the export of information technologies that can be used for the suppression of human rights, thus increasing the scope and scale of dual-use governance. In 2016 the focus was on the inclusion of so called 'cyber-surveillance technologies'[11], but recently AI applications such as facial recognition have also entered this debate as a possible next step for dual-use regulation.

Ad 6: Consumer protection

At this point, consumers do not receive adequate protection against unacceptable impact.[12] Many of these protections would need to be better covered in different existsing legislative instruments. As an example, the EU legal protection of consumers against unfair AI-driven personalized pricing is mostly principle-based, leading to uncertainty on its interpretation as long as there is no clarification in case law.[13]

Ad 7: Health

In the Medical Devices Regulation software is considered a medical device, making AI-driven applications in healthcare subject to certain requirements and a mandatory labelling scheme. Alignment of this regulation with a new legislative instrument on AI is important.

[10] Also: Directive 2009/103/EC on motor vehicle insurance; Directive 2007/46/EC on vehicle approval; Directive 2006/126/EC on requirements for driving licences
[11] Immenkamp, 2019
[12] BEUC, EU Rights for Consumers
[13] de Streel, Alexandre; Jacques, Florian (2019) : Personalised pricing and EU law, 30th European Conference of the International Telecommunications Society (ITS): "Towards a Connected and Automated Society", Helsinki, Finland, 16th-19th June, 2019, International Telecommunications Society (ITS), Calgary

Also, now that the Medical Devices Regulation has been postponed to enter into effect in May 2021, existing legislation, and the current practice of assessment and certification of AI-diven healthcare devices, should be critically looked at.

Ad 8: Energy & Environment

The High Level Expert group on AI (HLEG AI) has argued that sustainability and ecological responsibility of AI systems should be encouraged and has made "Societal and Environmental Well-being" one of the 7 requriements for Trustworthy AI of its Ethics Guidelines for Trustworthy AI.

According to the HLEG AI, it must be ensured that AI-systems operate in the most environmentally friendly way possible. The system's development, deployment and use process, as well as its entire supply chain, should be assessed in this regard, e.g. via a critical examination of the resource usage and energy consumption during training, opting for less harmful choices. Measures securing the environmental friendliness of AI systems' entire supply chain should be encouraged. This could trigger a necessery review and possible adaptation of EU engery regulations such as Directive 2010/30/EU on the indication by labelling and standard product information of the consumption of energy and other resources by energy-related products.

3. AI stress test for EU regulation

The foregoing calls for a much broader AI stress test for existing EU regulation. Three questions need to be answered in particular:

1. To what extent are the policy and legal objectives underlying these regulations affected by AI systems and in what ways?
2. What are the existing monitoring, information gathering and enforcement frameworks capable of providing meaningful and effective oversight to ensure that policy and legal objectives are still effectively achieved?
3. To what extent does existing legislation work in a way that they promote and safeguard the ethical principles and requirements (as described in the AI HLEG's Ethics Guidelines for Trustworthy AI)?

ALLAI points out that the AI HLEG has already advised the European Commission to perform this 'stress test' at EU level, in order to arrive at an unambiguous regulatory framework across Europe.

It should be noted that EU legislation does not affect all national legislative areas. Where the EU has no legislative competence, ALLAI advises the Dutch legislator to initiate the stress test at national level. Think of elements of labor law, social security law, administrative law and criminal law.

4. Basis for EU intervention

ALLAI fully agrees with the Inception Impact Assessment's reasoning that the objectives described cannot be reached effectively by Member States alone, but can be better reached at Union level. We agree that it is important that fragmantation would prevent the free circulation of goods and services containing AI and thus negatively affect the Digital Single Market. Fragmentation would also lead to divergence in levels of protection of citizens and society against the abovementioned harms and risks, which could lead to an unacceptable "race to the bottom" of AI regulation and protections in an effort to attract more AI investment.

B. Objectives and Policy Options

1. Objective

ALLAI supports the overall objective of the instrument, i.e. to ensure the development and uptake of lawful and trustworthy AI across the Single Market through the creation of an ecosystem of trust.

ALLAI would like to suggest the following additional aims of the initiative:

- To fill any legal lacunae either regarding the effectivenes, applicabilty or enforceability of existing EU law and where no EU law exists, so as to ensure that overall EU policy and legal objectives as regards trustworthy AI are promoted and safeguarded;
- In addition to what is expressed in aim (c) we strongly recommend to broaden this aim so as to include risks for people and society.
- As part of aim (c) and aim (e) it is important to be able to effectively monitor the future developments of AI within and outside the Single Market, so as to make sure that new opportunities for trustworthy AI are identified and promoted, but also that new challenges are adequately and timely addressed.
- In addition to aim (e), we suggest to set up a structure that includes not only the relevant authorities in the Member States, but also all relevant stakeholders, such as workers' and business' representatives, other civil society organisations, NGO's, academia (various disciplines), policy makers, etc.
- As an additional aim we reccommend to set up a European AI Authority as part of a global framework of AI Authorities. Such a framework could be set up as folows:
    - A global AI Authority;
    - Several regional sub-authorities (e.g. EU, the Americas, Asia, Oceanea, Middle East);
    - National executive authorities (either existing or new), for the EU to be appointed or set up by the Member States.

The roles and responsibilities of these authorities should be carefully considered, but they should have broad expertise of the different elements and impact domains of AI, including but not limited to technical, legal and ethical expertise, as well as knowledge of behavioural effects, labour market effects, economic and societal effects of AI.

2. Policy Options

On Option 0: "Baseline

ALLAI agrees that the current "Baseline" or Option "0" does not suffice to adequately address the risks and potential harms connected to AI and thus considers this option not viable.

On Option 1: "EU soft law"

The option of EU soft law should only apply to AI-systems or uses that have no adverse impact on people, society or the environment nor on our (fundamental) rights, democracy and the rule of law.

The two factor approach that was suggested by the European Commission in its Whitepaper on AI does however not suffice to determine these types of AI-systems. As an example consider targeted online advertising. The Commission will likely qualify advertising as a low-risk sector, and AI-driven targeted adds as a low risk AI-application. Targeted advertising however, has shown to have a potential segregating and dividing effect.

This is the reason why we recommend looking at the level of societal or personal impact of an AI-system or use to determine the risk level of the system.

On Option 2: "EU legislative instrument setting up a voluntary labelling scheme"

The labelling scheme is not a standalone option, but could be split into two options, the first to be joined with option 1 and second to be joined with option 3.

Where EU soft law would suffice (option 1), a voluntary labelling scheme could be an interesting addition for industry players to gain competitive advantage, or to feel confident to define or explore a niche area of application.

Where an EU legislative instrument establishing mandatory requirements for AI is necessary (option 3), a mandatory labelling scheme (like the CE-marking) could be considered to avoid any untrustworthy AI being deployed on the Digital Single Market. Current practices around certification of AI, such as those already executed in the healthcare sector, should be carefully reviewed, to see if those practices are sufficient, or should be amended or replaced by a new mandatory labelling scheme to avoid overlap with possible new labelling scheme.

Labels should not merely refer to the technical characteristics of the system, but more importantly also to the effects and impacts of the system.

On Option 3:"EU legislative instrument establishing mandatory requirements for AI"

ALLAI is in favour of introducing an EU legislative instrument, in the form of a combination of sub-options (a) and (b).

Exceptionally impactful AI

The following AI-systems or uses that are considered to be too impactful could give rise to the necessity of a ban, moratorium, strong restrictions or conditions for exceptional and controlled use:

- Indiscriminate use of facial recognition and other forms of biometric recognition either by state actors or by private actors;
- AI-powered mass surveillance (using facial/biometric recognition but also other forms of AI-tracking and/or identification such as through location services, online behaviour, etc.);
- Personal, physical or mental tracking, assessment, profiling, scoring and nudging through biometric and (online) behaviour recognition in violation of fundamental rights (AI-enabled Social/Citizen Scoring);
- Covert AI systems and deep fakes;
- Implanted human-AI interfaces;

Exceptional use of these technologies, such as for national security purposes or medical treatment or diagnosis, could be allowed but should be evidence based, necessary and proportionate and only be executed in controlled environments or cleary identified contexts and (if applicable) for limited periods of time.

As regards biometric recognition[14] (including facial recognition) we recommend that any use of biometric recognition only be allowed under the followoing cumulative conditions: i) there is a scientifically proven effect; (ii) it is used in controlled environment (e.g. a hospital); (iii) it is used under strict conditions (e.g. limited in time, for a specific purpose, etc.). Widespread and/or public use of AI-driven biometric recognition to surveil, trace, track, assess or categorise humans or human behaviour or emotions should not be allowed.

As regards AI-driven mass surveillance we refer to the recommendation of the HLEG AI in its Ethics Guidelines for Trustworthy AI that automatic identification raises strong concerns of both a legal and ethical nature, as it may have an unexpected impact on many psychological and sociocultural levels.

[14] Biometric recognition of micro-expressions, gait, (tone of) voice, heart rate, temperature, etc. is being used in various ways, one of which is to assess or even predict our behaviour, mental state, and emotions. As Barret et al. (*Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements*, 2019) have shown however, no sound scientific evidence exists to suggest that a person's inner emotions or mental state can be accurately 'read' from their facial expression, gait, heart rate, tone of voice or temperature, let alone that (future) behaviour could be predicted by it.

A proportionate use of control techniques in AI is needed to uphold the autonomy of European citizens. Clearly defining if, when and how AI can be used for mass surveillance both by public or private actors, differentiating between the identification of an individual versus the tracing and tracking of individual, will be crucial for the achievement of Trustworthy AI.

As regards AI-enabled Social Scoring we refer to the recommendation of the HLEG AI in its Ethics Guidelines for Trustworthy AI, that 'any form of citizen scoring should only be used if there is a clear justification, and where measures are proportionate and fair. Normative citizen scoring (general assessmentof "moral personality" or "ethical integrity") in all aspects and on a large scale by public authorities or private actors endangers these values, especially when used not in accordance with fundamental rights, and when used disproportionately and without a delineated and communicated legitimate purpose.' The HLEG AI also argues 'that citizen scoring – on a larger or smaller scale – is already often used in purely descriptive and domain-specific scorings (e.g. school systems, e-learning, and driver licences). Even in those more narrow applications, a fully transparent procedure should be made available to citizens, including information on the process, purpose and methodology of the scoring.' It also notes that 'mere transparency cannot prevent non-discrimination or ensure fairness, and is not the panacea against the problem of scoring and that ideally the possibility of opting out of the scoring mechanism without detriment should be provided – otherwise mechanisms for challenging and rectifying the scores must be given. This is particularly important in situations where an asymmetry of power exists between the parties. Such opt-out options should be ensured in the technology's design in circumstances where this is necessary to ensure compliance with fundamental rights and is necessary in a democratic society.'

As regards covert AI systems and deep fakes we again refer to the recommendation of the HLEG AI in its Ethics Guidelines for Trustworthy AI that 'human beings should always know if they are directly interacting with another human being or a machine, and it is the responsibility of AI practitioners that this is reliably achieved. AI practitioners should therefore ensure that humans are made aware of – or able to request and validate the fact that – they interact with an AI system (for instance, by issuing clear and transparent disclaimers). Note that borderline cases exist and complicate the matter (e.g. an AI-filtered voice spoken by a human). It should be borne in mind that the confusion between humans and machines could have multiple consequences such as attachment, influence, or reduction of the value of being human.'

As regards Implanted human-AI interfaces such as Elon Musk's Neuralink brain implant give rise to a plethora of legal (in particular as regards to (fundamental) rights that protect physical and psychological integrity, autonomy, free will, privacy etc.) and ethical concerns that should be carefully assessed and addressed in order to establish the appropriate regulatory framework.

Common denominators for determining the level of impact of AI

For all other AI applications, mandatory requirements on issues such as robustness, accuracy and reproducibility, traceability, transparency, human oversight and data governance[15] could be set, relative to the degree of impact of the AI-application.

We would like to stress again that these mandatory requirements should not be aimed merely at the data used to train and feed the AI-system, but also at the model(s) and algorithm(s) that comprise the system. These requirements should be met prior to the deployment of the system and be maintained throughout the use of the system.

We already expressed our concerns on the suggested 'two-factor' approach to identify 'high risk AI' (high risk application + high risk sector) that would then be subject to mandatory requirements. We also expressed our concerns on the list based approach of AI-applications that would be considered high risk 'as is', i.e. irrespective of the sector. For both structures, we fear that multiple AI applications or uses could fall (or actively be kept) outside the scope of the legislative initiative.[16]

In stead of the two-factor approach, we recommend to determine **common denominators for AI applications or uses that are to be considered high risk, or rather *medium or high impact*,** and would fall within the scope of the legal instrument. The following elements could serve as a guidance to set such common denominators:

- Common denominators should be determined based on the impact of the system on people and/or society, rather than (merely) on the technical characteristics of the system, or the sector in which the system is being used;
- 'AI impact' is to be considered both at individual and at societal/collective level whereas AI can impact both the individual as well as larger parts of our collective society;
- Context, severity, scale and likelihood of the impact is important to determine the appropriate and proportionate mandatory requirements, which could result in different requirements for different levels of impact;
- For high impact AI applications that generate unacceptable risks or pose threats of harm or systemic failure that are substantial, a precautionary and principle-based regulatory approach should be adopted;
- For medium impact AI applications a risk-based approach could be more appropriate.

On Option 4: "A combination of either or all of the previous options"

As can be concluded from the above, ALLAI would be most in favor of a combination of the policy options described in the Inception Impact Assessment. In short we would envisage the following structure:

- Soft law for low impact AI applications, including voluntary labelling
- EU instrument with mandatory labelling covering:

[15] We have elaborated on these requirements in our Final Analysis of the EU Whitepaper on AI, ALLAI (2020)
[16] Final Analysis of the EU Whitpaper on Artificial Intelligence, ALLAI (2020)

- Clear restrictions, conditions, safeguards and/or boundaries for a limited number of exceptionally impactful AI-applications or uses;
- Mandatory requirements for medium to high impact AI based on common denominators to determine the level of impact.

On: "Ex-ante and/or ex-post enforcement mechanisms"

While *ex-ante* and *ex-post* enforcement mechanisms are necessary, in our opinion, trustworthy AI primarily needs an *ex-durante* mechanism to ensure a continuous, systematic, socio-technical governance approach, looking at the technology from all perspectives and through various lenses.

This requires a multidisciplinary approach where policy makers, academics from a variety of fields (AI, data-science, law, ethics, philosophy, social sciences, psychology, economics, cyber security), social partners and NGO's work together on an ongoing basis.

It also requires socio-technical governance processes and structures that facilitate:
- Understanding of the potential legal, ethical and social effects of the AI-system and improve the governance/legislative choices based on that understanding;
- Monitoring of the workings and developments of AI to address and mitigate any new ill effects.

We thus call for a European AI Authority, as part of a global stucture of AI Authorities as mentioned above under B.1.

C. Preliminary Assessment of Expected Impacts

We agree that SME's should not be excluded from the application of the regulatory framework, precisely because of the high scalability of AI-systems.

As regards societal impact, we would like to draw attention to the fact that is still not clear whether AI will cause major loss of jobs or that this loss outweighs the number of new jobs it could bring.

The maintenance or acquisition of AI skills is necessary in order to allow people to adapt to the rapid developments in the field of AI is therefore important. But policy and financial resources will also need to be directed at education and skills development in areas that will not be threatened by AI systems (i.e. tasks in which human interaction is vital, such as public interest services related to health, safety and wellbeing of people and based on trust, where humans and machines cooperate, or tasks we would like human beings to continue doing).

As for environmental impacts, we expect that if energy regulation would be adapted to also cover AI, legislation could have a positive environmental impact.

∞

Authors

Virginia Dignum is professor of Artificial Intelligence at Umeå University, program director of the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS), co-founder of ALLAI, member of the European High Level Expert group on AI and of the World Economic Forum AI Board, and currently working as an expert advisor for UNICEF.

Catelijne Muller is co-founder and president of ALLAI, member of the European High Level Expert group on AI, Rapporteur on AI for the European Economic and Social Committee, and currently working as an expert advisor for the Council of Europe on AI & Human Rights, Democracy and the Rule of Law.

Andreas Theodorou is postdoctoral researcher on Responsible AI at Umeå University, member of the AI4EU consortium, member of the external ethics board of the ROXANNE project, and committee member on the IEEE Standards Association P70xx series of standards on AI.

ALLAI.