FINAL ANALYSIS OF THE EU WHITEPAPER ON AI

June 12, 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'. While the Whitepaper on AI of the European Commission does not propose legislation yet, it announces some bold legislative measures, that will likely materialize in the beginning of 2021.

By Virginia Dignum, Catelijne Muller and Andreas Theodorou[1]

We are pleased to see that the European Commission takes up many of the recommendations from a.o. the High-Level Expert Group on AI, the EESC and many other organisations, encouraging the uptake of AI technologies while also ensuring their compliance with European ethical norms, legal requirements and social values, underpinned by what it calls an "ecosystem of excellence and of trust".

Capitalising on AI in Europe – a forward-looking definition

As happy as we are to see a strong focus on governance of AI systems, we need to point out that the first step towards governance is to have a clear understanding of *what it is that needs governing* and *why*. The working definition of AI provided in the Whitepaper is: "AI is a collection of technologies that combine data, algorithms and computing power." This is later refined by claiming that AI is the combination of the first two, i.e. *data* and *algorithms*. This definition, however, applies to any piece of software ever written, not just AI. The governance of software in general, while being an important issue in itself, is beyond the scope of this paper.

So, what is AI? What makes AI different from other technologies so that specific governance and regulation is needed? Data and algorithms only refer to the ontological association of different components and not explicitly to the behaviour of the system. This could give organisations the opportunity to easily circumvent any regulation by claiming that a product is 'dumb' software and avoid compliance with any AI-specific requirements.

Even if we understand the desire to simplify things, AI is not simple, nor is its definition. The recent success of AI and the hype around it, have created a plethora of (pseudo) definitions, ranging from ultra-simplified ones as put forward in the Whitepaper, to pure magical ones. Depending on the focus and the context, AI has been referred to as: (i) a technology; (ii) a next step in digitization, by which the view is that everything is AI; (iii) a field of science aimed at modelling intelligence as means to understand intelligence; or (iv) a 'magic' tool or entity all-knowing, all-powerful, that happens to us without us having any power to control it.

[1] This paper is an addition to the 1st analysis of the EU Whitepaper on AI that ALLAI presented on February 9, 2020.

How we would describe it, AI technology is a piece of software with the following characteristics: it operates autonomously (i.e. without direct user control), its results are statistical (i.e. it does not link cause and effect) it is adaptable (i.e. it adapts its behaviour as it learns more about the context in which is applied) and it is interactive (i.e. its actions and results affect and are affected by us humans and our social and physical environment).

> AI systems are more than just the sum of their software components. AI systems also comprise the socio-technical system around it.

However, most importantly, AI systems are more than just the sum of their software components. AI systems also comprise the socio-technical system around it. When considering governance, 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.

It should also be noted that legal definitions (for the purpose of governance and regulation) differ from pure scientific definitions, whereas a number of different requirements must be met, such as inclusiveness, preciseness, permanence, comprehensiveness, and practicability. Some of these are legally binding requirements and some are considered good regulatory practice.

“AI Race” versus "AI Exploration"

As we have said before, the metaphor of AI as a race, promoted throughout the Whitepaper, is not only wrong but potentially even dangerous. A race implies a finish line and an explicit direction to follow. The idea of an “ultimate algorithm or an AI-ruler” simply feeds into the unscientific narrative of ‘superintelligence’, damaging public trust in the technology and distracting from real-world governance problems. The field of AI is vast and its full potential is far from being fully explored.

Even though we are now seeing many results from the application of a specific type of technique (deep learning, which is roughly based on artificial neural networks), one just needs to look at the past to know that it is unwise to put all your eggs in the same basket. Deep learning applications are far from being intelligent enough to solve all our problems. Even if such systems would excel at identifying patterns, e.g. identify cats in pictures, or cancer cells, with an accuracy close to or higher than humans, the system will have no understanding of the meaning of cat, or a cancer cell. It will only be able to provide a label to a specific pattern. And even then, it will still have great difficulties in describing the properties of a cat, let alone that it be able to use its understanding of cats to understand dogs, or chickens.

> Ultimately, trustworthy AI cannot be a choice between an accurate black box AI-system or an explainable but less accurate AI-system. We need both.

Ultimately, trustworthy AI cannot be a choice between an accurate black box AI-system or an explainable but less accurate AI-system. We need both. This means that a new generation of AI systems is needed that integrate data-driven approaches with knowledge-driven, reasoning-based approaches, with human values and principles in the centre. Here European research has an important advantage: since the early AI days, European researchers have excelled in a variety of approaches to design and verify artificially intelligent systems. Rather than blindly racing with others, European researchers approach the problem as explorers: mapping a wide field of possibilities and plotting promising results, such as in symbolic AI (that can link cause to effect) and hybrid systems.

The Whitepaper, in a one-liner statement, acknowledges the need for such systems for the purposes of *explainability*. But the advantages of hybrid systems go beyond explainability. They entail the ability to speed-up and/or restrain learning, validate and verify the machine learning model, and more. The Commission needs to acknowledge leading European research efforts in this direction, and encourage these approaches. By equating data as *the* essential component for AI, the White Paper excludes non-data-driven approaches, e.g. expert systems, knowledge reasoning and representation, reactive planning, argumentation and others, from being considered AI and, therefore, from being subject to the regulatory framework. It also misses the opportunity to promote trustworthy AI as a competitive advantage and attach the right incentives to foster and drive trustworthy AI.

Bias, transparency, and the need for socio-technical software engineering

The focus on data-driven systems extends even further where the Whitepaper focuses on bias in relation to data. Luckily it dismisses the often heard argument that both humans and artefacts can act on bias, by stating that intelligent systems can enshrine and further disseminate and amplify our biases while also obscuring their existence and without the social control mechanisms that govern human behaviour.

> The design of any artefact is in itself an accumulation of biased choices, ranging from the inputs considered to the goals set to optimize for.

It however overlooks that not all biases are the result of low-quality data. The design of any artefact is in itself an accumulation of biased choices, ranging from the inputs considered to the goals set to optimize for; Does the system optimize for pure efficiency, or does it take the effect on workers and the environment into account? Is the goal of the system to find as many potential fraudsters as possible, or does it avoid flagging innocent people? All these choices are in one way or another driven by the inherent biases of the person(s) making them.

In short, suggesting that we can remove all biases in (or even with) AI is wishful thinking at best and an error of language at worst. In either case, for the purposes of any regulatory framework we should not merely focus on technical solutions at dataset 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 biases transparent; and
c) continuously monitor the workings of the systems to mitigate the ill effects of any biases.

To this effect, the Whitepaper correctly promotes the need for traceability on the decisions made by the human actors related to the design, development, and deployment of a system.

This form of transparency within the social structure helps users (both expert and non-expert users) to calibrate their trust to the machine, testers to debug the system, auditors to investigate incidents and determine accountability and liability.

These are all existing approaches in software engineering that we can use, in stead of reinvent for AI. The Whitepaper unfortunately does not acknowledge this broader perspective on transparency by merely focusing on a binary dogma of "opaque high-performing systems versus transparent low-performing systems". As such it promotes transparency only for expert technical users and not for the broader group of non-technical deployers, users and those affected.

Bringing all forces together

We welcome the effort to address the fragmented AI landscape in Europe by bringing together AI researchers, focusing on SMEs and partnering with the private and public sectors. In addition, the we would recommend the following to bring all forces together: (i) foster multidisciplinarity in research, by involving other disciplines such as law, ethics, philosophy, psychology, labour sciences, humanities, the economy, etc.; (ii) involve relevant stakeholders (trade unions, business organisations, consumer organisations, NGOs) in the debate on AI, but also as equal partners in EU-funded research and other projects such as the Public Private Partnership on AI, the sector dialogues, the *Adopt AI* programme and the *lighthouse centre*; and (iii) continue to educate the broader public on the opportunities and challenges of AI.

AI does not operate in a lawless world

The Whitepaper acknowledges the fact that AI does not operate in a lawless world, thus ending the discussion on whether AI should be regulated or not and (hopefully) silencing the voices that claim that AI is an unregulated technology (and should stay that way). Secondly, it emphasizes that AI has impact on our fundamental rights. This is important, because many of us take our fundamental rights and freedoms for granted.

> Our freedom of speech and expression, our right to a private life, our right to a fair trial, to fair and open elections, to assembly and demonstration and our right not to be discriminated against (...) these are the rights that are jeopardized by certain types and uses of AI.

Our freedom of speech and expression, our right to a private life, our right to a fair trial, to fair and open elections, to assembly and demonstration and our right not to be discriminated against, these are all rights that are simply part of our lives. But these are also the rights that are jeopardized by certain types and uses of AI. For example, facial recognition has already shown to affect our right to freedom of assembly and demonstration when people in Hong Kong started covering their faces and using lasers to avoid being caught by facial recognition cameras. For this reason, the Council of Europe, the 'house 'of the European Convention on Human Rights and the European Court of Human Rights, is currently investigating a binding legal instrument for AI.

Liability

We welcome the announced adjustments to the existing safety and liability regimes. the Commission correctly takes a clear stance on the applicability of existing liability regimes to AI. We also welcome announcement that the Commission will build on those regimes to address the new risks AI can create, tackle enforcement lacunas where it is difficult to determine the actual responsible economic operator and make them adaptable to the changing functionality of AI systems. Persons having suffered harm as a result of an AI-system, should have the same level of protection and means of redress as persons having suffered harm from any other tool, according to the Whitepaper.

We would like to reitterate that we continue to oppose the introduction of any form of legal personality for AI. This would hollow out the preventive remedial effect of liability law and poses a serious risk of moral hazard in both the development and use of AI, where it creates opportunities for abuse.

Regulating AI

Beyond the existing legal framework (that i.e. ensures consumer protection, addresses unfair commercial practices, protects personal data and privacy and sets rules for specific sectors), the Commission explores the need for new mandatory requirements for AI. Before we go into these requirements, we however want to comment on the approach the Commission is exploring, i.e. that (only) "high-risk AI" would need to comply with these requirements.

The 'two-factor' approach for high-risk AI

According to the Whitepaper, two cumulative elements constitute high-risk AI: (i) a high-risk sector and (ii) a high-risk use of the AI application. Only if these two requirements are met, the system would be subject to new mandatory requirements. The Whitepaper hurries to add that there might be exceptional instances, where the use of an AI application is considered a high-risk *as is*, i.e. in recruitment and when worker's rights are impacted. It also qualifies biometric recognition a high-risk application, irrespective of the sector in which it is used.

The list of high-risk sectors is to be exhaustive (whilst periodically reviewed) and the Commission already indicates the following sectors as potentially high-risk: healthcare, transport, energy and parts of the public sector. The second criterion, being that the AI application is used in a risky manner, is more loosely defined, suggesting that different risk-levels could be considered based on the level of impact on the rights of an individual or a company. We would suggest to add society and the environment here.

While this looks reasonable at first sight (not each an every AI application or use constitutes a high risk) the chosen system might still have some gaps. We agree that and AI-application used to channel the open WIFI-signal in a mall would not be considered a high risk, while that same AI-application used in a military setting or a hospital would (for cyber security or privacy reasons).

But what if we look at the opposite situation? Following the logic of the Whitepaper, an AI application (high- or low-risk) used in *a low-risk* sector would in principle not be subject to mandatory requirements.

> At this point we think that the 'high-risk sector-requirement' might not the most effective way to achieve what the Commission wants to achieve, which is to avoid that any and all AI needs to be subjected to strict regulation.

As a thought experiment, let us consider targeted advertising, search engines and movie recommender systems. The Commission will likely qualify advertising, information and entertainment as low-risk sectors, while targeted advertising has shown to have a potential segregating and dividing effect, search engines have shown to make biased search predictions and video recommender systems prioritize 'likes' over quality and diverse content amplifying fake news and disturbing footage. If we were to address these particular undesirable effects of AI, they would all have to count as 'exceptions' and be subject to mandatory requirements.

Secondly, working with lists, either for high-risk sectors or exceptional AI applications or uses that would be considered high-risk '*as is*') is always suboptimal.

Lists are never exhaustive. They need amendments as the state of the art develops over time, new risks might emerge and new insights will rise.

While the two factor approach could be maintained, but need to be further explored first, we recommend to additionally drawing up common characteristics of AI applications or uses that are to be considered high risk '*as is*', irrespective of the sector in which they are used. We will contribute towards this in any way we can.

Mandatory Requirements

The mandatory requirements that Commission proposes are robustness, accuracy, reproducibility, transparency, human oversight and data governance. We agree on the need for these new mandatory requirements and provide additional insights in how to shape these below.[2]

But first and foremost, we would like to stredd taht 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.

Robustness, accuracy and reproducibility

Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm.

This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured.

In particular, AI-systems should be resilient to attack, malicious use and abuse resulting from e.g. hacking, data poisoning or model leakage. For AI systems to be considered secure, possible unintended applications of the AI system (e.g. dual-use applications) and potential abuse of the system by malicious actors should be taken into account, and steps should be taken to prevent and mitigate these. They should have a fall back plan for unintended and harmful situations and safety of the systems should be tested continuously and pro-actively.

AI-systems should be accurate in their predictions and make correct judgements. Whenever this is occasionally impossible, the systems should be able to indicate how likely the inaccuracy is, and a maximum acceptable level of inaccuracy (if any) should be determined. Apart from the need for reproducibility of the results of AI systems, they should be reliable. The system should work properly with a wide range of inputs and in a wide range of situations.

[2] See also: Ethics Guidelines for Trustworthy AI, 2019

Transparency

Transparency should include traceability, explainability and open communication. This means that processes need to be transparent, the use, capabilities and purpose of AI systems openly communicated, and decisions explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested. An explanation as to why an AI-system has generated a particular output or decision (and what combination of input factors contributed to that) is currently not always possible. These cases are referred to as 'blackboxes' and require special attention. Other explainability measures (e.g. traceability, auditability and transparent communication on system capabilities) may provide sufficient counterbalance, however, the AI-system and the outcomes needs to respect fundamental rights and ethical principles.

Human oversight

Human oversight helps ensuring that an AI system does not undermine human autonomy or causes other adverse effects. Oversight may be achieved through governance mechanisms such a s a human-in-the- loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach. HITL refers to the capability for human intervention in every decision cycle of the system. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. These mechanisms should include the option not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system.

Data governance

AI systems must guarantee privacy and data protection throughout a system's entire lifecycle. This includes the information initially provided by the user, as well as the information generated about the user over the course of their interaction with the system. Digital records of data may allow AI systems to infer not only individuals' preferences, but also their sexual orientation, age, gender, religious or political views. Data collected about individuals and information inferred, may not be used to unlawfully or unfairly discriminate against them.

When data is gathered, it may contain (historical) biases, inaccuracies, errors and mistakes. In addition, feeding malicious data into an AI system may change its behaviour, particularly with self-learning systems. Data sets should thus be tested, cleaned, optimized and documented at each step (planning, training, testing and deployment). This should also apply to AI systems that were not developed in-house but acquired elsewhere.

Socio-technical and 'human-in-command' approach in stead of prior conformity assessment

A truly eye-catching announcement is the idea of putting in place *prior conformity assessments* for high-risk AI that would need to go through a rigorous testing and validation processes before entering the EU internal market. The Commission explicitly mentions that this obligation will apply to all actors irrespective of their location.

While we acknowledge need for conformity testing of AI and the relevance of all the requirements, we fear that a one-off (or even a regularly repeated) conformity assessment will not suffice to guarantee the trustworthy and human-centric development, deployment and use of AI in a sustainable manner.

In our opinion, trustworthy AI needs a continuous, systematic socio-technical approach, looking at the technology from all perspectives and through various lenses. For policy making, 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, businesses and NGO's work together on an ongoing basis. For organisational strategy, it requires involvement of all levels of an organisation, from management to compliance and from front to back office, are involved in the process on an ongoing basis.

In this sense, the Whitepaper has a slightly 'fatalistic' flavour to it, where it seems to think that AI 'overcomes us', leaving us no other option than to regulate its use. But we do have other options. One of which is the option to decide not to accept a certain type of AI(-use) at all, for example because it will create a world that we do not want to live in. This is what we have been calling the 'human-in-command' approach to AI that we need to foster.

Biometric Recognition

The Whitepaper states that biometric recognition ((tone of) voice, gait, temperature, heartrate, blood pressure, skin color, odor, and facial features) is already heavily restricted by the GDPR and that there are specific risks for human rights. It also opens the discussion on if, and if so, under what conditions to allow biometric recognition.

> *"Let's pause and figure out if there are any situations, and if so, under what circumstances facial recognition should be authorised" (...) "As it stands right now, GDPR would say 'don't use it', because you cannot get consent."*
> - Margarethe Vestager, VP European Commission

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.

It should also be noted that the GDPR only restricts the processing of biometric data to some extent. The GDPR defines biometric data as "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". Many biometric recognition technologies, however, are not designed to uniquely identify a person, but only to assess a person's behaviour or emotions. These uses might not fall under the definition of biometric data (processing) under the GDPR.

AI-driven biometric recognition also affects our broader right to respect for private life, identity, autonomy and psychological integrity by creating a situation in which we are (constantly) being watched, followed and identified. This could have a psychological 'chilling effect', where people might feel inclined to adapt their behaviour to a certain norm. This constitutes an invasion of our fundamental right to privacy (moral and psychological integrity). Furthermore, AI-driven biometric recognition could affect other fundamental rights and freedoms, such as freedom of assembly and the right not to be discriminated against.

We recommend that any use of biometric recognition only be allowed under the follwoing 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.

The Impact of AI on work and skills

We note that the Whitepaper lacks a strategy on how to address the impact of AI on work, whereas this was an explicit element of the 2018 European Strategy on Artificial Intelligence.

We advocate 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

We would like to draw special attention to AI used in hiring, firing and worker assessment and evaluation processes. The White Paper 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.

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. 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).

Education in living and working with AI systems will be required for all, beginning at an early age, in order to ensure that people retain autonomy and control in their work and lives. Education about ethics, law and fundamental rights in relation to AI is particularly important, since AI has a significant impact in these areas.

AI and Corona

While the Whitepaper was published before the Corona-crisis hit Europe, we nevertheless to take this opportunity to address the topic of AI to combat or gain a better understanding of the Corona-virus and COVID-19 and its consequences.

AI can contribute to gaining a better understanding of the Corona-virus and COVID-19 and their societal and economic consequences. It has the potential to help protect people from exposure, help find a vaccine or explore treatment options. The responsible thing for AI researchers and professionals to do at this moment is to put their expertise to use for the analysis and understanding of the spread of the coronavirus, the possible effects of policies, the long term impact on society and economy, and other such questions. However, in order to do this responsibly, it is important to be open and clear about what AI can and cannot do. As we often stress, AI is not magic nor is it a solution to all our problems. Some of the main requirements for responsible development and use of AI include robustness, transparency and respect for human rights and ethical principles.

Firstly systems need to be robust. We need to look carefully to the techniques and approaches used. In particular, the use of data-driven methods to forecast the spread of the coronavirus, is potentially problematic. These methods 'learn' by correlating data from the past and at this moment we just don't have enough data about similar situations.

Results from the past are no guarantee for the future, especially when the future looks to be so very different from anything we knew from the past. Moreover, the little data that is available is incomplete and biased. For example, it is certain that there are many more infected persons that those that have been confirmed, and there is not a complete accounting of those that have recovered. Using the existing data for machine learning approaches can lead to many false negatives and false positives.

The current situation may be a case for which model-driven methods are more suitable than data-driven ones. Indeed there are several groups at this moment working on (agent-based) simulations. However, the selection of models needs to be grounded on sound research in epidemiology, sociology and psychology, together with the suitable computational representations. This type of models tends to be sensitive to design assumptions and initial parameters. Before proposing these systems to support policy makers, sensitivity analysis tests need to be conducted. From a research perspective, work on hybrid approaches that combine data-driven and model-driven are especially important to be working on at the moment.

In either case, data-driven or model-driven, transparency is paramount. Which models and datasets have been used and why, which experts have been consulted, how have the system been tested and evaluated. Without explicit answers to these questions, results cannot be trusted. At this moment more than ever, the world cannot afford to take decisions based on 'black boxes'. If your system cannot provide explanation of its results, please think twice before considering using it to provide any analysis or forecast on the pandemics. At the same time, efforts to use AI to support identifying fake news and limit its spread are highly recommended.

Tracking and tracing and health-monitoring apps

According to virologists and epidemeologists, opening the society and the economy from lockdown requires efficient tracking, tracing, monitoring and protecting of people's health. Currently, many apps are being developed with the hopes of performing activities that have usually (and historically) been done by professionals. Worldwide, many governments have placed a large amount of trust in tracking and tracing apps as a means of opening up societies again.

The deployment of these kinds of apps is a very radical step. It is therefore important to critically examine the usefulness, necessity and effectiveness of the apps, as well as their societal and legal impact, before a decision is made to use them. There must still be the option of not using the apps, and less invasive solutions should be prioritised.

The effectiveness and reliability of tracking and tracing apps is extremely important, because ineffectiveness and unreliability can lead to many false positives and false negatives, a false sense of security, and thus a greater risk of contamination. Initial scientific simulations raise serious doubts as to whether a tracking app will have any positive effect on the spread of the virus at all, even with 80% or 90% use. Achieving these high percentages will be very challenging in and of itself, considering the number of people with inferior digital skills, people without a mobile phone, and children. Also, an app cannot register specific circumstances, such as the presence of plexiglass and windows or wearing of personal protective equipment. It is even uncertain whether a technology exists that can reliably and accurately measure the 'proximity' of mobile phones.

Moreover, these apps lead to the (partial) setting aside of various human rights and freedoms, as they touch on our freedom of association, right to safety, right to non-discrimination, and right to privacy.

While very important, privacy is about much more than our personal data and anonymity. Privacy is also about the right not to be followed, tracked, and put under surveillance. It has been scientifically proven that when people know they are being followed, they start to behave differently. According to the European Court of Human Rights, this 'chilling effect' is an invasion of our privacy. The same broad concept of privacy should be included in the AI debate.

There is a risk that data collected (now or in the future) will not only be used to fight the current pandemic, but also to profile, categorise and score people for different purposes. In the more distant future it is even possible to imagine that "function creep" could lead to unwanted types of profiling in supervision and surveillance, acceptance for insurance or social benefits, hiring or dismissal, etc. The data collected using such apps may therefore under no circumstances be used for profiling, risk scoring, classification, or prediction.

Moreover, any AI solution deployed under these extraordinary circumstances and even with the best of intentions, will set a precedent, whether we like it or not. Previous crises have shown that, despite every good intention, such measures will in practice never go away.

The use of AI during this pandemic should thus always be measured and weighed against several considerations, such as: (i) is it effective and reliable? (ii) do less invasive solutions exist? (iii) do its benefits outweigh societal, ethical and fundamental rights concerns? and (iv) can a responsible trade-off be achieved between conflicting fundamental rights and freedoms? Moreover, these kinds of systems may not be deployed under any form of obligation or coercion.

We urge policy-makers not to succumb to techno-solutionism too readily. Given the gravity of the situation, we recommend that applications linked to projects designed to help control the pandemic be grounded in sound research in epidemiology, sociology, psychology, law, ethics and systems sciences. Before deciding on the use of these systems, efficacy, necessity and sensitivity analysis and simulations need to be conducted.

We therefore call on the European Commission to assume a leadership role to ensure better coordination within Europe in terms of applied AI solutions and approaches used to understand and fight the Coronavirus pandemic and its consequences. Such leadership should encompass the application of common ethical standards, consolidation of resources, as well as ensuring the transparency, openness and comparability of collected data to empower further research, development and scalability of promising AI solutions.

Conclusions

The White Paper is the European Commission's first concrete attempt at discussing AI policy beyond the high-level statements of previous Communications. In this sense, the Commission takes up a rule setting role (rather than a referee role). In our opinion, this is a good first step.

"In a game with no rules, nobody wins." - Virginia Dignum

If we were to draw the analogy with a game, independently of who is playing the game, without rules no one wins. Moreover, the potential impact of AI both positive as negative is too large to be left outside democratic oversight. While the ideas of the Commission need further elaboration and depth, the true the leap forward would be not only to focus on "Trustworthy AI made in Europe" as an alternative to AI made by the existing tech giants, but to promote trustworthy AI as a competitive advantage and incentivize and invest in the institutions, research and frameworks that can set this new AI playing field.

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.

ALLAI refers to Stichting ALLAI Nederland, a foundation under Dutch Law. No entity or person connected to ALLAI, including its Board Members, Advisory Board Members, employees, experts, volunteers and agents, is responsible or liable for any direct or indirect loss or damage suffered by any person or entity relying wholly or partially on this communication.

Herengracht 247
1016 BH Amsterdam
welkom@allai.nl
www.allai.nl