EUROPEAN COMMISSION
Directorate-General for Communications Networks, Technology

The High-Level Expert Group on Artificial Intelligence

Outcomes of the AI HLEG Workshop of
20 September 2018
Brussels

TABLE OF CONTENT

INTRODUCTION 1

OUTCOMES FROM THE BREAKOUT SESSIONS 2

1. Trusted AI 2
2. Transparency & Accountability 7
3. Industry & Ecosystems Uptake of AI 10
4. Use Cases for the Guidelines 14
5. AI Infrastructure and Enablers 17

CONCLUSIONS 21

INTRODUCTION

On 20 September 2018, the High-Level Expert Group on ("AI HLEG") held its first workshop in Brussels. Participants of the workshop included the members of the AI HLEG (experts and observers), members of the AI HLEG reserve list, a limited number of ad hoc experts and Commission officials.

The purpose of the workshop was to dive in-depth into the themes that the AI HLEG will address in its two deliverables, the draft AI ethics guidelines and the AI policy and investment recommendations, and gather input from the workshop participants to advance the drafting process.

The workshop started with a plenary session. After a brief introduction by Lucilla Sioli, Director of Artificial Intelligence and Digital Industries at DG Connect, the Chair of the AI HLEG – Pekka Ala-Pietilä – welcomed the workshop participants and set out the timeline and work process ahead. The two Vice-Chairs of the AI HLEG, Nozha Boujemaa and Barry O'Sullivan, subsequently explained the current draft structure of the two deliverables.

For the Guidelines, the draft structure focused on the concept of "Trusted AI", which is the overall result to be achieved by ensuring that the ethical intent is in place when dealing with

AI, but also that – besides the intent – trusted AI is also implemented in the technology and its use. For the Recommendations, the draft structure focused on the impacts or goals that the AI HLEG would like Europe to achieve on the one hand, and the different enablers that are needed to ensure such impact on the other.

After the plenary session, the workshop participants were split into four groups and spread over four different rooms for the breakout sessions. Five themes were discussed in more details over the day – one overarching theme, and two per Deliverable:

1) The concept of Trusted AI (discussed by all four groups – Overarching)
2) Transparency and Accountability (discussed by two groups – Deliverable 1)
3) Industry and Ecosystems Uptake of AI (discussed by two groups – Deliverable 2)
4) Use Cases for the Guidelines (discussed by two groups – Deliverable 1)
5) AI Enablers and Infrastructure (discussed by two groups – Deliverable 2)

In each group, a moderator and a rapporteur volunteered. The moderator's task was to guide the discussion and try to steer the group towards an answer on the guiding questions that were formulated for each theme. The rapporteur's task was to take note of the discussion and report the outcomes thereof to the rest of the group at the end of the day.

In what follows, the outcomes of the different breakout sessions have been summarised based on the reports of the rapporteurs who volunteered and based on the notes taken by the Commission officials assigned thereto.

OUTCOMES FROM THE BREAKOUT SESSIONS

1. Trusted AI

a) Introduction

"Trusted AI" is a cross-cutting theme in the work of the AI HLEG. It is both the goal which the AI HLEG aims to achieve through the draft AI ethics guidelines (i.e. the first deliverable), as well as a necessary part of the second deliverable, focusing on AI policy & investment recommendations.

In Europe, an important tradition of human values and fundamental rights exist, whereby technology, business and ethics go hand in hand. The development and uptake of AI is inextricably intertwined with the ability to trust in AI. Without such trust, consumers and users more generally will shy away from adopting AI even if beneficial, which could potentially lead to a loss of the tremendous benefits that AI can generate. Trusted AI is achieved not merely through regulation, but also by putting in place a human-oriented and ethical mind-set by those dealing with AI, in each stage of the process.

A common understanding of the concept "Trusted AI" and of the concepts and actions that lay behind it/are intertwined therewith, is crucial to understand how it can be achieved. Trusted AI therefore envisages both (1) the right ethical intent when dealing with AI, and (2) the correct implementation thereof. While the first meeting of the AI HLEG already touched upon the core values that need to be considered in an AI context (intent), this breakout session allowed for a discussion on the implementation of those values (implementation).

The **guiding questions for the discussion** were the following:

- What does it mean to ensure trusted – i.e. responsible and robust – AI?
- Which tools can be implemented to achieve such trust?
- What does it mean to have trusted AI that supports and enables European businesses to be successful in applying AI?
- How can the above tools be implemented to achieve such businesses' success?

b) Outcomes from the breakout sessions

What is AI?

Some experts mentioned the need for a **classification** of AI systems at the outset, and identified four types of AI which could be in need of specific recommendations:

- Internet AI
- Business AI
- Sense AI (vision AI, facial recognition)
- Autonomous system / trusted system

What is Trusted AI?

The discussion highlighted the fact that Trusted AI is a **broad notion** that refers to many different aspects, which can imply different things in different contexts. Nevertheless, it was agreed that some **common principles** can be distinguished which form a general baseline.

It was mentioned that it would be useful to have a **list of topics** or aspects that define what Trusted AI should be like. Moreover, AI could be broken down in separate items to become concrete and secure trust in its different applications.

Some experts expressed a preference for the term 'trustworthy', instead of 'trusted', because it is in our hands to demonstrate positive outcomes of AI applications which will generate trust. Some stated that Trust is about the application of AI, not about AI itself, since AI is but a tool.

Some stated that the **foundational pillars/elements** to make Trusted AI work are transparency, accountability, safety, intent and sustainability; auditing (could be at various levels: institutional/company/third party) or verification that the requirements are met (in terms of security, privacy...); and documentation transparency adapted to the various types of users (professional / workers…). It was also mentioned that Trustability turns around the issue of deception, i.e. the line between reality and that what appears to the user.

It was stressed that transparency and explainability play a major role in enabling the user to understand and trust AI applications. **Trust comes from understanding**, and a lack of understanding translates into a lack of trust, that is why it is important to invest in educating the general public about what AI is and what it can do, to increase trust of future generations. Some stated that it should be the government's responsibility to made citizens aware and knowledgeable about AI. This also to ensure that AI is used when it can be beneficial. For instance, in healthcare where human lives are at stake, citizens need to trust AI systems before being able/confident to provide data. Medical teams need to understand and to rely on AI in

the first place, and citizens should be educated about the use of their healthcare data to increase trust. People need to know how their data can help the development of AI systems to serve better their needs. Knowledge enhancing is thus not only necessary to counter the risks of AI, but also to **make citizens aware of the benefits thereof**.

However, it can be asked how much knowledge of how an AI system works users should have in order to trust it. Often, the concept of trusted AI suggests that all aspects of how AI works should be explained, but we should not ask more of machines than we do of doctors. Nevertheless, this also requires safety guarantees. Users trust systems that guarantee safety, but at the moment there are no guarantees that data-driven systems are trustworthy, and therefore it is very important to develop such guarantees and thus enable people to trust AI systems, even without understanding in detail their inner workings.

A number of experts agreed that making AI trustworthy and responsible refers not just to the development stage, but to the entire lifetime of an application, which also includes deployment and use. AI applications are very continuous, i.e. they are living services the outcomes of which can change in time depending on changes in inputs, and therefore there is a need for lifetime trust. Furthermore, trust in technology refers to the behaviour of the technology in question, but also to the makers of that technology. Trusted AI will ultimately be about trusted entities and organisations. Companies should be supported by providing them with a list of aspects that brings more precision. At the same time, companies should also be held accountable for their systems, based on their declarations as to the functions that their systems maximise.

Some experts proposed the following methodology to address the concept of Trusted AI:

a. Define the high level concept: what is trust?
b. Cover the Outside view: how to present it? Document the important properties of the system/service, adapt the message to the audience or type of users. Clarify and simplify the interface between the AI system and the users.
c. Cover the Inside view: What function should the AI system deliver, what are the requirements and how are they implemented.

To enable trust, some common principles could be defined, but there was agreement that they must be developed/implemented/translated for each use-case.

Available tools?

Existing algorithmic assessment tools should be considered in order to make an AI system trustable, to show how data is collected, processed or used, taking into account steps such as verification, validity, security, control, unbiased data. Tools that can be used include guidelines for different types of industries, identification of bias and missing data, corrected data, transparency and methodology, accuracy measurement, auditing systems, standards.

Increasingly, also new tools are available to help companies check if their technology is trustable and explainable. Such tools help with thinking about trust from the very beginning, not just in the form of checks at the end of the development process. Those tools reflect the need for a systematic, comprehensive approach in order to ensure that AI applications can be trusted. Such a **systematic approach** should cover not just the skews that may affect the

application, but also the processes involved. AI applications are often developed with the participation of many **external partners**, and therefore the involvement of such partners has to be taken into account as well when discussing the notion of trusted AI and who is ultimately responsible.

Open standards would be needed and trusted third party institutions could be used to validate/certify them (bearing in mind that too early standardization might hamper innovation). Trust could also be enabled by measures ranging from a certification label to an AI-like "Hippocrates" oath.

Next to accuracy measures, we also need to have **confidence measures**, and be transparent on when the system can know something with confidence or not. In addition, it was stated that, often, looking at the data is easier than looking at the system, and may thus be the way to go.

It was acknowledged that we need methods to properly **translate the values** that we wish to ensure **into the design of the AI tool**. Translations are contextual, and determined by the people who are translating it (typically code developers) – the latter not always being aware of how to make such translations or not having clear instructions thereon.

It was stated that to ensure trust, the process will in any case need to be **adaptive.** One particular specificity of AI based systems is the **dynamicity**, so it is not sufficient to guarantee that the requirements for trust are met at the launch of the system; the corresponding performances should be maintained and verified over time, e.g. through life-cycle management. We should thus **not only build trust but also maintain trust**.

Communication is at the core of trust building: it is important to adapt the message to the various types of users and inform them accurately about: the intend of system, what benefit it offers them, its accuracy, potential bias, how complete is algorithm, how adaptable it is to unforeseen situation, how robust to attack, how many IP does it capture, what impact it might have on the user, how decisions are made, etc. The various **stakeholders are not trained to understand AI,** people need to have the skills to carry out audit, understand the potential gap between the implementation and the guidelines. In this **multidisciplinary context**, it is key to establish dialogue and common understanding between lawyers, policy makers and technicians, each having its "own language".

As a template or analogy to trusted AI, we can also look at the **tools we use to foster trust in humans**, these primarily being laws, standards and certifications. Moreover, a major tool to ensure trusted AI can be AI itself: interactive normativity was raised in this regard. AI can be used to partly solve ethics, by turning ethics into a data science problem.

Businesses face the following issue: they want to deliver to their customer a trustable AI tool, and want to be able to show the trustability of the system. It is however difficult to prove to the customer that the contract is respected (i.e. delivery of a trustable tool). In many cases, such proof will also be needed for the acceptability of the system by showing that the AI tool is more trustworthy than a "human" tool/human decision.

The way transparency is guaranteed should also take into account the business model and not hamper competitiveness in disclosing secrets. An audit by a trusted third party, possibly imposed by the regulation, could be a solution, providing trust in the solution.

An attempt was made to **categorize the different tools available** to reach trusted AI, covering (i) **technical tools** (of which different types exist) and (ii) **governance tools** (e.g. ethics guidelines) are the most important ones. Each company has its own production and approval processes, and the fundamental question concerns whether AI's particularities are just an extra step in the existing processes of a company to consider, or whether it is something entirely separate that requires its own validation process. Some stated that, ideally, we ensure that the process/tools for trusted AI fit into an already existing process, which not only facilitates the exercise but also maximizes the chance that this will be applied.

Finally, it was stressed that **remedies and redress** should be in place and the impact of problems should be assessed. Trust is increased if there is a line of defense (e.g. ability to turn to the Ombudsman).

Seal / Certification of Trusted AI?

There was mention of the idea of developing a **seal of trusted AI**, reflecting the requirement that systems comply with certain standards. Nevertheless, in many cases such certification would not be easy, especially given that 95% of current AI is software and therefore 'invisible'. Such systems may require tests that check for transparency, explainability, etc. Another complication is the need to ensure that such certification does not hurt companies, and instead is helping them by making their goods and services more valuable because they are known to be trustworthy. Europe has high values and that could be a competitive advantage because **European companies are seen as the most trustworthy companions** for consumers.

Policy & Regulation

Compliance with policy and regulations, as well as values, is at the core of trusted AI. Both algorithmic and data aspects should be monitored and regulated, but the first step is to check which regulations apply and whether they are sufficient. Some experts believe that no new regulation is required at all, but only further development building on what is already in place. **Legislation gaps** should however be identified. In any case, AI must be compatible with existing EU legislation and Human rights (e.g. Fair trial, effective remedy…), as well as the GDPR. The discussion included thus not only ethics, but also the law and legal accountability.

There is also a risk that **regulation may be premature**, given the speed at which technology is evolving. It is, however, necessary to develop use cases. The IEEE standards relating to AI are a good starting point for reaching consensus on what it means to be non-biased, transparent, fair, etc. Indeed, **standards based on robust methodology and a use-cases** / operational approach are a good way to go. At the same time, it is important to ensure that **trusted AI helps European businesses** and does not obstruct them, especially given the intense competition with the US and China.

Regulation of AI also has to take into account the fact that AI is a **dual-use technology**. For example, facial recognition can be used in emergency situations to reunite families, but it can also be used to stamp out political dissent. Furthermore, the line between AI and regular software is blurry, which raises the question whether there should be different regulations for different techniques. Some experts believe that AI should not be treated differently.

Moreover, it should be acknowledged that there are no perfect data sets: data sets inevitably are incomplete and can contain errors.

Research

Some mentioned that the fact that AI is based on learning, doesn't fundamentally change the game. It is but a system that optimizes parameters in a certain way. Methods to ensure trusted AI already exist, but **research is needed** to identify why the optimization went one way or another. Many research programs on were launched and this is fundamental for us if we wish to understand why AI reached a certain decision when based on deep learning techniques. Europe should thus be in the game also as concerns **research in explainable AI**.

Inclusiveness

It is necessary to ensure that persons with disabilities are part of the conversation, because **trust matters to everybody**. Persons with disabilities are usually early adopters of new technologies, but unfortunately, there has been a negative trend with respect to accessibility. Twenty years ago, persons with disabilities were able to operate most standard appliances, such as microwave ovens and fridges, but that is no longer the case. It is also important to make aspects related to impact on people more clear and visible in the structure of the second Deliverable of the group (policy and investment recommendations).

In addition, in order to ensure responsible AI, **all relevant stakeholders should be involved** in the discussion. Importantly, fairness should be ensured not only towards the customer/user of the AI tool, but also towards society at large.

Finally, it should be borne in mind that, sometimes, a particular AI application can be great from the point of view of increased personalization for the user, while eroding some social paths or having an impact on collective dimensions of society that we do not immediately think of.

2. Transparency & Accountability

a) Introduction

Transparency is one of the key principles mentioned when discussing the ethics of AI. Rather than being a value in itself, transparency can act as a tool to ensure that the fundamental values of our society are respected. The concept therefore hinges closely together with accountability, as well as with other pertinent concepts such as auditability, fairness, explicability, traceability, non-discrimination and interpretability.

In order to apprehend how transparency and the other mentioned concepts can be used as a tool to implement and achieve Trusted AI, a scoping and disentangling of these different concepts is warranted. The **guiding questions for the discussion** on Transparency and Accountability were the following:

- What do the concepts transparency and accountability mean in an AI-context?
- How can their concrete and practical operationalization be ensured?

b) Outcomes from the breakout sessions

Refinement of concepts

At the outset, it was pointed out that there is a need for refinement of concepts. We need to be clear about whether transparency is a value or whether it should be seen only as a tool. One needs to be aware that there is no static "right approach". The only pragmatic and feasible approach is much more **dynamic** and can take into account future research and societies, as well as different and new human reactions. We will therefore need to adapt our ethical framework on an ongoing basis.

When speaking of AI, some stated that it should be preferable to use the terms "explainability" and "interpretability", and that "Transparency" should rather be used for the development process of AI. A special case for transparency by companies developing AI was singled out. Such companies should be transparent about the production process (biases in data) and not only to inform/advertise about the capabilities of their AI products/services. Factsheets with a set of questions to answer, could make AI systems more trusted.

Further, several experts expressed the opinion that the work of the AI HLEG should not focus on defining transparency and accountability as phenomena, but to single out which of their aspects/dimensions are **relevant to AI**. Transparency could have different degrees/scales depending on who is the addressee. Moreover, it was pointed out that sometimes, in certain contexts, a certain extent of opacity may be needed.

Another approach to transparency dimensions was demonstrated by giving the example of the black box - what is inside a black box requires different level of transparency. Four aspects of **accountability** were singled out in this regard:

- Input: focus on quality of data
- Throughput: focus on design of algorithms
- Output: focus on decision
- Outcome: focus on impacts – negative/positive consequences

Some of the experts advocated a pragmatic and realistic approach focusing on what is at stake, moderating expectations by reducing the complexity of the topic to a subset of AI systems. The same principle of a non-generalist approach was also put forward for the future selection of use cases: it is unlikely that all use cases can be fit into one single model. Instead, several principles should be singled out in order to see how they can be applied, by taking into account a mix of values and ways to achieve them. The no-one-fit-all model approach is for instance applied in the context of impact assessments. Indeed, different impact assessments should be used: Value based AI, Privacy and security AI and so on.

Responsibility & Accountability

Experts singled out as one of the unique aspects of AI the phenomena of **shared responsibilities**, referring to the example of the PPPs. Here the question concerns the boundaries between public and private sectors. There is a need for clarification on what is government responsibility for algorithmic decision-making and what are the responsibilities

of the private sector. It was pointed out that algorithmic accountability should be subject to public scrutiny and consultation.

Regarding the phenomenon of complex responsibility, experts put forward that citizens and machines are different from a legal point of view. In the reality of shared responsibility, with many actors involved, from a legal point of view it is important to identify the type of responsibility (i.e. civil or criminal responsibility). This raises the question whether new standards should be introduced, or whether the existing ones are enough.

As a reflection on this topic, some experts referred to the human rights paradigm; in order for AI to be accountable, it should follow **human rights legislation**. The GDPR already tackles many of the issues, but an **analysis of legal gaps**, where additional regulation mat be needed, is deemed useful.

The reigning opinion was however that there **should not be any AI-specific regulation**, as our normal legal systems should be general enough to cover the particular case of AI as well.

Three levels of accountability were singled out:

1. micro level – people, researchers and professionals
2. mezzo level – public institutions, companies, universities
3. macro level – how politicians are accountable to citizens

There should be clear **responsibility allocation**: who are the responsible persons behind the system and possibly liable? Here one should take into account that there are vulnerable groups. e.g. for disabled persons, who are extensive users of AI systems, it should be important to understand why a system does not work, and whom to contact in case of a failure issue. For example a web page that isn't accessible for a person with impaired sight? Who fixes it? Who is legally responsible?

Communication with citizens

Experts agreed that there should be a special **focus on communication with citizens** to ensure their buy in of AI. In fact, not only humans should be in the loop, but **also society should be in the loop**. To enhance transparency there is a need for the involvement of citizens; they should be able to co-construct AI processes, which in turn would enhance transparency. Practices quoted in this regard were the use of surveys to ask for people's opinion and organizing of workshops to explain how AI is functioning.

Also, AI systems should be easy to use. One should avoid information overload and stick to the concept of **balanced information**. To make systems accountable, a simplification of the system is useful. Convenience and safety of use also matter.

Some mentioned that there should be a ban on 'social hallucination', this being explained as the importance of not deceiving the users: there should be clarity who is behind the screen - AI or a person? AI should never hide its identity. There should be a universal capture “I am not a human”.

While there cannot always be a human (final) intervention, **if there is an impact on people they should have a say.** A legal framework is needed to ensure consumer trust. It is

important to educate consumers, to ensure that not everything needs to be explained every time. This also relates to nudging.

Finally, experts pointed out that transparency is also about explainability of systems. To be accountable, AI should be simplified, certified and audited.

Other points that were mentioned:

- To ensure transparency, the data sets must be accessible for verification and challenging. We **should be able to challenge the data algorithmically**. There should be algorithmic tools for control and to provide truth of trust. To ensure transparency we should inspect at each stage of AI development and use: what data is used; how data is used; and how the algorithm is used. Generally, we must ensure that, if the AI systems fail, they fail gracefully.
- Algorithmic accountability matters. Here **hybrid accountability** is an important concept discussed in the academy. Transparency relates not only to the level of single decision making but also to structural biases. Users may not be thinking of long term consequences. Not every single person can take into account everything.
- AI systems with existential results on people's life (police/ judicial systems decisions) should be singled out as a specific case, and extra attention should be paid thereto.
- Some experts pointed out the strong focus in the discussions on the bad aspects of AI. We have to remember that AI can also be used for good, and that most companies want to do good. Hence, we should also **pay attention to the positive aspects**.
- Finally, there is a need to **streamline the different ongoing initiatives**. There are many initiatives from the European commission, and different expert groups were established. The AI HLEG should coordinate therewith to avoid risk of overlap.

3. Industry & Ecosystems Uptake of AI

a) Introduction

One of the main goals of the AI policy and Investment Recommendations is creating an impact by enabling the uptake of AI. Industry and Ecosystems are crucial stakeholders in this regard. The aim of the breakout session was therefore to explore what precisely the envisaged impact on those stakeholders should entail, and which measures can be taken to achieve such impact.

The **guiding questions for the discussion** were the following:

- How can AI's uptake be accelerated across the entire European industry?
- Which measures can we take to ensure a level playing field in AI's uptake across all sectors, companies of all sizes, and all regions?

b) Outcomes from the breakout sessions

Part of the discussions explored the main aspects that AI uptake involves, including human-centred approaches, democratisation and inclusion, and broad usage. There was emphasis on the importance of understanding both the potential and the limitations of AI, and on the need to have the capacity to adopt AI (e.g. the necessary skills to develop and/or implement AI solutions).

Furthermore, it was stressed that **uptake has many different dimensions**, such as products, processes, business models, and training. Ultimately, the notion of uptake should cover use (by industry), empowerment (by the governance ecosystem), and user friendliness (e.g. acceptance/trust, especially in healthcare, and especially by consumers).

The discussions also included mentions of the current situation in the EU, where the vast majority of companies are only starting to explore AI. The **categories that need more help** are start-ups and SMEs, which are actually the larger groups among European businesses. Such entities need more assistance compared with others who already adopted AI and now they only need to scale up.

To accelerate/encourage the uptake of AI in Europe, it was suggested to adopt an **ecosystem approach** that provides support in parallel both for the EU AI industry and for the use of AI across all sectors/industries. The idea of **domain-specific ecosystems was mentioned**, and as a start, there could be first an experimental ecosystem, for example in health or automotive. An EU network of science excellence hubs could also have a significant contribution, as would tailored uptake, more entrepreneurs and funding, education (defined broadly, including knowledge transfer), a start-up culture, data exchange across markets (e.g. health data).

The current investment plans in Europe were described as significant, but still very low-scale compared to China, for example. This raised the question whether AI should really be promoted everywhere across the economy, or whether it is better to select one sector or a few sectors to focus on, in particular sectors that are critical for the welfare or survival of citizens (e.g. health). Similarly, should all companies be encouraged and supported in adopting AI, including very small SMEs? The funding planned will not be sufficient to support everyone.

Some stated that the EU should start where we are already strong, i.e. build around existing major assets, mostly in the industrial sector. It would be necessary to **set priorities** based on areas that are high-potential or critical. In this context, ethics could also be an enabler, not an obstacle, and it is a European strength. There is also a need for **more AI infrastructure** in Europe, because the future of AI is also infrastructure. At the moment, the US has both the necessary experts and all the infrastructure, and companies in Europe find it difficult to hire AI **talent** because the ones coming from the US are too expensive and the ones trained in Europe have a different skillset.

European AI story

Europe should also insist more on the **European AI story**, which is very important and should not be compared to the US and China. The story of European AI was described as largely untold. Europe has a larger number of research papers on AI and it is a powerhouse of research. We have to be more aware of this aspect, and tell a stronger European story to the

rest of the world. At the same time, there should be awareness that Europe is a leader in science but that **research is not sufficiently used to generate business**. Several experts insisted on the importance of **investment** for enabling **hospitals and doctors** to use AI. Another priority sector should be **public administration**. EU institutions and national governments have a responsibility to invest in AI for public goods such as **healthcare**.

Experts representing start-ups mentioned the importance of **accessing the market**, in the case of a small company such as a start-up. In the area of AI, such companies usually have corporate customers, which makes this area primarily B2B. The importance of **venture capital** was also emphasized, including the difficulties arising from the fact that after the first obstacle, which is to convince venture capital firms to invest, such investors want to sell after ten years, and the buyer is usually one of the major players. When the company is sold to one of those major players outside Europe, the team usually leaves Europe too. **Corporate venture investment** would be a better solution, because it would keep talent inside the company. Other ideas mentioned included a **fund** created with money committed by companies and used to buy start-ups, and devising an **exit strategy** that would be only for Europe (given that there are venture capital investors that invest exclusively in Europe). Another problem is the fact that the market in Europe is very consolidated, and start-ups often have to start somewhere else before doing business in Europe.

Importance of Data

Some of the key enablers for accelerating the uptake of AI are **data interoperability** and portability, complemented by the **free flow of data and open data**. For examples, tech titans in the US have signed an agreement on data in healthcare, to achieve interoperability. Getting the data in digital form is one of the most important aspects. Publishing data may not necessarily result in a level-playing field, and might benefit big companies more than smaller ones. However, even though large companies may have a lot of data, they usually also need other types of data from others.

The recommendation was that the **data ecosystem be strengthened** in Europe, including by defining **regulatory sandboxes**. Public data should be open, and small companies should be given access to build on public data. Health data should be allowed to flow across European markets, and public administrations should be enabled to share public data. Current legislation has elements that obstruct the use of data, and sandboxing would be useful to explore where the problems are, instead of inventing something new. Level-playing field considerations are also important with respect to algorithms, and there is a need for independent solutions in order to avoid having to enter somebody else's ecosystem without any control. At the same time, AI should be regarded as part of the larger picture of digital transformation, and the broader landscape of data collection and curation.

Skills & Education

It is very important to remove obstacles. Big companies invest heavily in AI, but medium-sized companies find it harder. Even though they do see the benefits, they do not have suitable staff and may not even know what they can use AI for. That is why education is key, and why we need centers where AI is made available and where there are people who know at least a

little about AI. It was suggested that the European Commission invest in education for new talent and retraining existing workers. There is not enough time to wait for new generations, so it is very important to **invest in retraining** and invest as much as possible in developing AI capacity quickly. Up-skilling of domain experts for AI could play a major role. In this context, there was also mention of a possible European driving license in AI, similar to the European Computer Driving Licence (ECDL). Some problems do not require super deep researchers, and can be solved with basic methods.

However, education systems are national, and it is difficult to do more at the European level. Companies could play a larger role, in order to make sure that people affected by the AI transformation are retrained or engage in new business ventures. **Co-funding from companies** and the EU at the same time would be very important for skills and their impact as the great accelerator. **Free online courses** have also been demonstrated to be of great interest to existing employees interested in re and up-skilling.

Companies are also often confronted with the fact that new employees arrive without the necessary skills. That is why cooperation between universities and industry is very important. Such cooperation is very important also with respect to the mobility of experts between universities and industry. It is important to retain AI leaders, otherwise European universities and industry will become weaker.

The speed of the uptake of AI could also be increased by stimulating acceptance of AI among CEOs and by creating a **digital single AI market**. Free courses in AI were described as one example of best practice for attracting more people to AI.

Other points mentioned

- Should competition rules be changed so as to prevent big companies from buying the smaller ones? In this context, size is not all that matters. It is also important to look at the data held by a company and at how important that data is. Furthermore, some AI technologies should perhaps be labeled as sensitive (e.g. weapons and surveillance technology), and investment by foreigners made subject to investment controls.
- Other possible specific actions mentioned included the **mapping of sectors** according to AI potential and readiness, and the **mapping of existing initiatives** (with a view to finding gaps, and to replicating geographically what works). It was suggested that the Commission select the health and automotive **ecosystems for a 360-degree round** on what they need to scale with AI. The Commission could also support user groups by industry and a **fund linking research and industry**. It would also be important to carry out an analysis/survey of the cost of errors in decision-making, with a view to balancing the cost of mistakes and the impact.
- Several experts suggested that the CLAIRE and ELLIS initiatives should be merged.

4. Use Cases for the Guidelines

a) Introduction

The Guidelines' purpose is not to list a high level set of ethical principles to be considered when using AI, but rather to provide concrete and operational guidance for those dealing with AI on how to ensure an ethical approach thereto.

Ethical AI is not an abstract or theoretical concept, given that is always related to the particular case, application and context in which it is used. Against that background, the Guidelines will comprise a set of Use Cases to clarify how Trusted AI can be implemented by following a comprehensive check-list, that can be applied by analogy to other Use-Cases.

The guiding questions for the discussion were the following:

- Which use-cases are most useful to consider when preparing Draft AI Ethics Guidelines for Trusted AI?
- Which use-cases can be used to challenge/test the draft Guidelines, ensuring that the document adequately serves its purpose?

b) Outcomes from the breakout sessions

Methodology to identify use cases

As a first step, the need to create a **methodology** for selecting use cases was identified. In the context of establishing such methodology, the following points were mentioned:

- to select and discuss use-cases and out of this discussion to crystallize principles and not vice versa, and then:
- to have a combination of real (then to be anonymized) and fictional cases . Some members disagreed and stressed that there is no need for real cases, but that the guidelines should rather treat well known examples/stereotypical cases, or real cases that are at least anonymized and generalized when it concerns public and private organizations.
- to have cases with possible positive and negative applications – including ambiguous cases
- to have cases where EU has advantageous position
- to have cases not focused on EU position but to be representative for real problems
- to define how many use-cases should be prepared

The following questions were likewise identified to design a methodology for use-case selection: Should the scope include best-case and worst-case scenarios? What would be the testing principles? Is the goal to test the use of AI?

Some stated that the AI HLEG should work on ambiguous cases, and not only on straightforward cases (for instance, a use-case which also implies "job losses" could be meaningful). It was found that the Guidelines would be even more robust if they also address more apparently innocuous decisions that are delegated to machines but that create important ethical matters because of scale of deployment of the decision-making. The selected cases should consider questions going beyond principles and that are rather borderline.

The guidelines could be selected based on their risk, with a good mix of high and normal risks, and should not cover what is already covered (e.g. GDPR).

An ethical analysis/ethics assessment should be a preliminary step of the methodology, and the line between ethics & responsibility should be borne in mind.

The existing formal framework to describe all use cases implemented with ISO could be taken into account, and a couple of domains were identified: fintech, manufacturing, healthcare, ICT, legal education, digital marketing, maintenance and support, social infrastructure, security and defence, work and life.

It was suggested that the group should check whether the guidelines as applied to the use cases result in outcomes that we generally expect in a moral compass. Moreover, it was noted that the checklist(s) should be comprehensive: without concrete examples, it is not comprehensive and remains abstract.

The handling of contradictory positives should also be part of the guidelines, as in some situations there are more than one positive aspects which may contradict each other (e.g. in healthcare, a choice may need to be made between saving the economy money or prolonging someone's life etc.).

As concerns the audience of guidelines, it was stated that both the public and private sector should feel that they are addressed by the guidelines.

Finally, it was noted that identifying use-cases is also a good tool to assess potential policy gaps.

Use Cases Identification

In a second step, the experts discussed **suggestions for use-cases**, the list of which can be found here below:

- Social support assistants: Electronic butler, Companion robot, Conversation robot, digital inclusion – chatbots
- Health care topics (Early diagnosis systems, Alzheimer disease/deterioration in patients condition)
- Justice / law-enforcement decisions (Persistence and fairness of judgement, collecting case law - finding precedents)
- Smart cities
- AI for human augmentation (Healthcare)
- Health Care Agent / robo-nurses
- Early disease detection/prevention agent/ Suicide prevention Agent / disabilities agent
- Organ donation
- Human/worker collaboration and Human/Machine interactions.
- Manufacturing
- Scoring people (Transparency of ranking systems) / Risk prediction
- Liability on smart home environment and control systems
- Self-driving cars (safety & liability issue)
- Call center agent

- predictive algorithms; changes in education and learning
- Stock exchange market
- Killing robots
- Risk assessment systems / Decision taking systems
- consumer rights – digital assistants; credit scoring
- social robots; data paid by taxpayers / algorithms for decision making
- B2B and industry applications;
- education;
- politics;
- social robots – healthcare; smart environment, smart cities
- Law (e.g. contract – two parties, liability management, law making)
- Integrated border management and smart borders (benefits those people who are privileged, identify vulnerable group, rapidly evolving area and demands attention)
- Recruitment process (possible discrimination and bias produced by AI possible even if in the end the decision is made by HR manager)
- Public procurement (What are criteria behind the selection? Does it include assessment of price and quality? Verify how it is implemented in different structures)
- Biometric data in seamless services (Could we use systems (e.g. phones) regardless and without providing biometric data?)
- Children sex robots (use by pedophiles)
- Voice Recognition for elderly and disabled people
- Pattern recognition for elderly and disabled people (Website in accessible way to use it with alternative means of assistive technology)
- Face recognition
- Image recognition (visually impaired: hearing what is described in the picture)

Autonomous weapon systems were discussed fiercely as there were pros and cons amongst the experts. Some noted that this point is a (too) complex matter which should potentially be discussed by specialists in an ad hoc working group.

Horizontal themes:

Finally, some cross-cutting/horizontal themes were identified:

- Data Sharing:
    - To which extent is it moral **not to** share data (e.g. rare diseases)? Not sharing your data may in some instances risk other people's life.

- Health data:
    - The collection of data for health purposes has been frequently in the news, particularly when it went wrong (e.g. when patients didn't give consent for their data to be used). Committees were set up to look at ethical implications when this happens within national health system. The solution could be legislation/convention agreements to prevent such a case and to set up rules for such use of personal data, to the extent not yet sufficiently covered by the GDPR.

- In some Member States, researchers have access to public health data. Such data could be of great help for people to remain healthy but in wrong hands, it could have a devastating effect (e.g. insurance company). Some mentioned that there is currently a huge leeway to use data for scientific purposes.
- It was noted that ownership of data is not only about data protection. We need to find balance between the opportunity of AI and open data and respect for privacy: the establishment of data banks / cooperatives, respecting data protection laws, could feed useful AI applications.

- Positive agenda of AI:
  - Some stated that there is no need to talk about what is wrong with algorithmic decision-making in itself but rather how such decision-making is used. Developing the ethics guidelines should not occur in a defensive way only.
  - We need to also set a positive agenda of AI. Also when looking at the future of work for instance, health and safety of workers can be increased (e.g. exoskeleton). There should be no unwarranted ex ante judgment.
  - In this regard, "in-between" zones could be identified: e.g. digital marketing which is legal, yet applied in political context to influence opinions. While this is not illegal strictly speaking, it does have an impact on democracy – such cases should be particularly looked into.

- Regarding the consumer field:
  - The question was raised on the risk of categorization of consumers, in a certain category or receiving an individualized profile as consumers, which may have an impact on the offers/access that they may receive (e.g. Credit)
  - The atomization of the demand side could make it impossible to compare prices properly and thus to make good consumer choices

- Other cases that were discussed:
  - Smart homes: it was found that liability should be more properly defined to cover the questions raised in such scenario.
  - Online trading and cartels: it was asked whether our competition rules are still able to cope with algorithmic manipulation.
  - Moral responsibility of companies/stakeholders and for instance suicide prevention: If a company detects or is aware of a person about to commit suicide, what would be the responsibility as a company? Is it your responsibility to save a life or to not look into a person's messages and ignore that?

5. AI Infrastructure and Enablers

a) Introduction

The Recommendations will focus on a number of goals in relation to AI in Europe, such as (i) creating an impact on the Business side (ensuring the uptake of AI by industry and

ecosystems), creating an impact on the Public side (ensuring the uptake of AI by the public sector, and in the Public-to-Citizens market), or creating World Class Research capabilities in Europe.

In order to achieve the envisaged goals, a number of enablers are available in Europe, such as for instance funding and investment, infrastructure, skills and human capital development. This breakout session therefore aimed to identify the different enablers of AI, and to discuss how such enablers can be ensured.

The guiding questions for the discussion were the following:

- What are the most important building blocks at EU level to enable European businesses to successfully apply AI?
- How can such building blocks be ensured, and how can investments therein be stimulated?

b) Outcomes from the breakout sessions

To identify the relevant building blocks necessary for successful adoption of AI across the European economy, it would be necessary to first know what were are precisely aiming towards.

Priorities?

A discussion may be warranted on specific potential benefits in specific sectors, followed by a decision on how to achieve those benefits. Education and health were mentioned as areas that should be treated as priorities for AI-related investment, as it was found important to invest in the basics, in things that are essential to citizens, first. Some raised that another priority could be tourism, as tourism experiences are strong in Europe.

Europe should build on EU strengths such as robotics and other AI areas. Europe already has some results to build on, but they are not sufficiently exploited. The EU should act fast in the area of AI in general, because other regions talk in months not years, and speed is critical for European competitiveness. A grass roots approach would be appropriate, whereby it is necessary to act at both EU and Member State level. The EU should also think bigger, and consider large-scale moonshot projects.

General building blocks

Specific examples of building blocks for a successful AI-enabled European economy included data, skills and computing resources. The latter category included hardware as well, which is critical but tends to get overlooked. Hardware is more than computing. It was stated by some experts that Europe today does not have the expertise anymore to build computers, and that it is also dependent on AI-specific chips, even though it has beaten the US and China in robotics.

The Internet was another example of crucial infrastructure element and essential building block. For example, if everybody were to use a self-driving car today, the current Internet infrastructure would not be able to cope with the load. Europe would need serious infrastructure, and it also needs to think in cross-border terms in that respect. There should be common European facilities, i.e. physical infrastructure, where top-class experts could meet

and collaborate, but also share their knowledge with the new generations and with other researchers in Europe.

Companies considering the adoption of AI solutions need access to testing facilities, to be able to test a solution before committing to considerable investments. One example of a platform offering such facilities is TeraLab. There should also be zones of protected technology development, and there should be common accessible infrastructure with easy access for SMEs. There should be HPC resources tailored to AI, and the EUROHPC project was described as an important opportunity. Other important elements included high speed networks, ubiquitous and robust communications, and 5G (in particular high speed between large R&D centers and testing facilities).

Data

Data sharing was described as a very important building block, but there was also emphasis on the need to keep the data private where necessary. Furthermore, in many cases a company's existence depends on data. In such cases, sharing data is actually sharing money. However, sharing access to knowledge could be very beneficial also for the source company, which would be able to monetize its knowledge. Companies should be encouraged to engage in such sharing of access to knowledge. Software infrastructure solutions for such knowledge sharing already exist. For example, researchers can use clusters in the cloud to run algorithms. It would mean offering not the data itself, but only services accessing the aggregate result. The data would stay with the author, but third parties could run algorithms. The data sharing would take place via APIs. Europe could fund the development of such system. One similar example mentioned was the OPAL ('Open Algorithms') project/platform. There was also emphasis on the need to distinguish training data sets from the marketplace, and on the importance of determining which sectors should be the focus. Data sets could be developed by sector, and there could also be a CERN-like centre for data.

Data sharing between companies could also take place through consortia or syndicated platforms. Companies in different segments of the same sector, for example dairy producers and cattle farmers in agriculture, could decide together to pool their data and generate greater benefit for everybody involved. Europe should also encourage industry to share more with academia in specific areas or on specific issues. Industry has many use cases and problems that are interesting to tackle. There was mention of a smart city laboratory, for public good, generating data available to all and allowing any company to test solutions in that environment. CERN was another example. It generates data that is made available in a systematic manner.

There were also suggestions to develop EU reference data platforms, which should be high performance and competitive and used by all, and to use such platforms to generate data in particular in areas where the EU can lead. Major sources of data mentioned included CERN data (e.g. used for automation systems, for the car industry), automotive data, smart city data (e.g. environment, traffic management), and industrial data spaces. Satellite data is available to all, but it is underused in Europe. It is necessary to find the right context to exploit such data.

Another idea suggested is the use of digital twinning of physical systems for simulation and for generating data (IoT, smart grids, smart cities, environment, etc., with IoT systems used to collect data, and data used to train AI systems). Platforms would be needed in order to make European scientific excellent results easily deployable in any company, and thus provide a competitive offer compared to GAFA platforms (that some of Europe's prestigious car companies are using). Bringing research results to the market faster was deemed critical. There is a need to ensure that research translates easily into products.

Talent

Access to talent, including not only scientists but also integrators, managers, or for adapting organisations and processes in industry, was described as a top priority. AI will impact every single area where software is used. The possible solutions envisaged included up-skilling programmes, training the trainers programmes, building curricula in cooperation with industry, picking talent at the international level (e.g. at international conferences), scalable training accessible to all and adapted to the audience (MOOCs for basic level), and fostering grassroots skills for data engineers, data mining experts, topic matter experts, and software experts (who could be upgraded in 4 months).

There should be investment in excellence in AI, and in networking the top scientists, and such excellence should then be made available to all. Salaries should also be competitive. There should also be a focus on training for young people.

There was a suggestion that the EU could, for example, provide 1:1 top-up funding to schools spending money on AI-related skills (i.e. 1 euro provide by the EU for each euro spent by a school on such skills). Another similar idea was about existing models of private and university co-funding of PhDs and Masters degrees, which could be used as inspiration for similar EU co-funding. There were also mentions of examples of online courses such as an online Masters degree in Spain, very good for lifelong learning, and other mechanisms that could support the development of skills for the future, including incentives in the form of tax credits and tax breaks, for companies that providing training in advanced digital skills for their employees. It was also important to enable people to move back and forth between academia and industry, for example by creating more interaction programmes between academia and industry (supported at EU level), and to educate people locally. Industry should be more involved in training people.

Funding

Funding was described as a major enabler, and there was emphasis on the need for easier and simpler access to funding. Such access should be faster and more agile, and there should be an environment that facilitates the growth of start-ups and the development of a culture of risk and innovation. There is a need for a culture of empowerment, because – as someone noted – at the moment the best thing for a European start-up is to be bought by a US company.

Regulation

Smart regulation was mentioned as another important building block. AI is not unregulated. There were mentions of existing legislation, including the safety and liability frameworks and fundamental rights. Smart regulation was described as looking at what is already in place and

determining what else needs to be added. Data gathering should be combined with sandboxing, to be used as a testing environment for identifying where the problems and the obstacles are.

There should be regulatory sandboxes and EU shared facilities available for development and testing. For businesses, it is highly important to have legal certainty on AI deployment. The question of what happens after release on the market and who is liable is very important for the management of a company. At the same time, consumers and users need to be able to trust the products. However, it was stressed that the EU should be careful not to over-regulate, take into account regulation which already exists, and only adopt new rules where certain use cases require it.

There should be pan-European tools and frameworks/jurisdiction supporting the Digital Single Market and addressing fragmentation, such as pan-European certification, pan-European governance clauses, and pan-European transparency rules. The GDPR was mentioned as a good example to build on, but some experts found that the GDPR had created fragmentation and that now certifications may need to follow 28 different directives, hence their suggestion for a pan-European certification governance model.

Ecosystems

There was a specific suggestion related to the idea of developing industrial/sectoral ecosystems and finding the enablers (i.e., funding, regulation, data, etc.) for such ecosystems to become reality. The specific suggestion was to start now a project around which an ecosystem with suitable enablers could develop. Such a project could begin with three sectors, and then be replicated to other areas. The first three sectors mentioned were factory of the future (industry in a day / industry 4.0), health (focus on a sub-topic) and automotive/mobility.

Communication

It was noted that the EU should drastically improve its communication. One suggestion was to develop an 'excellent AI made in the EU' trademark. The EU should also increase awareness and knowledge of best use-case examples. Showing that AI solutions work is essential for adoption on a large scale. Finally, the discussions on how to support the development and use of AI should not ignore what Europe can do using AI also outside Europe.

CONCLUSIONS

At the end of the workshop, the Rapporteurs of each breakout session reported back to the group in plenary setting. The sessions generated lively discussions, and their outcomes constitute an important input that can be used when drafting the different sections of the AI HLEG deliverables, together with the feedback gathered through the European AI Alliance.

The Chair of the AI HLEG thanked the workshop participants, and – based on the discussions in the breakout sessions – formulated the idea to start setting up AI ecosystems in a small number of sectors that could to be prioritized. These ecosystems could run as a pilot project, and could be scaled up / extended to other sectors if successful. The idea put forward entails a

practical approach to start testing out the uptake of AI in parallel with the drafting process of the recommendations, so that the outcomes and lessons learned can be taken on board therein.

As a next step, the AI HLEG met on 8 and 9 October in Helsinki, in the margin of the AI Forum, where it continued its work on the different themes of the deliverables in sub-groups. A next meeting is planned on the 8th of November 2018 in Brussels. The next AI HLEG workshop will take place on 13 December.

The current timeline for delivery of the outputs is as follows:

- For deliverable 1 (the Guidelines), by the end of the year, a first draft should be made available for public consultation, and a final version should be presented around March 2019.
- As concerns the Recommendations, a final version should be presented around May 2019, yet the AI HLEG is currently aiming to deliver a first draft thereof also by the end of the year, to allow a public consultation simultaneously with the Guidelines.

The timeline is ambitious in light of the work ahead, but the motivation is strong to work on something meaningful together.