CLAIRE

Response to the European Commission White Paper -
On Artificial Intelligence
- A European approach to
excellence and trust

12 June 2020

The Hague, NL

Content

Foreword

This document is the official response of the Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE; see also http://www.claire-ai.org) to the European Commission's White Paper "On Artificial Intelligence – A European approach to excellence and trust", dated 19 February 2020.[1] It was drafted by a group of experts from CLAIRE selected by the CLAIRE leadership ("task force white paper response") and subsequently circulated for comments to CLAIRE's International Advisory Board and the CLAIRE membership at large. Based on the comments received, the task force produced this final version, which was subsequently officially endorsed by the Board of the CLAIRE Association.

The task force consisted of Prof. Fredrik Heintz, Linköping University, Sweden (co-chair); Prof. Jeroen van den Hoven, TU Delft, The Netherlands (co-chair); Prof. Holger Hoos, Leiden University, The Netherlands; Prof. Ana Paiva, University of Lisbon, Portugal; Dr. Josef Urban, Czech Technical University, Czech Republic; Prof. Bart Verheij, University of Groningen, The Netherlands; and Anna Tahovská, Czech Technical University, Czech Republic (support).

The members of CLAIRE's International Advisory Board are Prof. Noriko Arai, National Institute of Informatics, Japan; Dr. Frédérick Bordry, CERN, Switzerland; Prof. Alan Mackworth, University of British Columbia, Canada; Dr. Francesca Rossi, IBM Research, USA; Prof. Robert Jan-Smits, TU Eindhoven, The Netherlands; Prof. Manuela Veloso, J.P. Morgan Research and Carnegie Mellon University, USA; Prof. Wolfgang Wahlster, German Research Center for Artificial Intelligence and Saarland University, Germany; and Prof. Toby Walsh, University of New South Wales, Australia.

The views and recommendations expressed in this document are based on the CLAIRE vision (see https://claire-ai.org/claire-vision) that is supported by the 375 groups and organisations that form the CLAIRE Research Network, the 3506 individual supporters of the CLAIRE vision, and the governments of nine European member states that have officially confirmed their support for the CLAIRE vision (Belgium, Czech Republic, Finland, Greece, Italy, Luxembourg, Netherlands, Slovak Republic and Spain).

---

[1] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Key recommendations

In response to the white paper, much of which aligns well with the CLAIRE vision for European excellence in human-centred AI (https://claire-ai.org/claire-vision), CLAIRE makes the following key recommendations to the European Commission:

1. **Make sure to complement the push for AI regulation with swift and substantial investment into AI research**, including curiosity-driven, foundational research - Europe cannot be a leader in AI regulation without being a leader in AI, and it cannot be a leader in AI applications or innovations without being a leader in foundational AI research.

2. **Create streamlined allocation mechanisms of AI research support**, focussing on those researchers and institutions with a track record of excellence in AI as well as on those with demonstrated potential for excellence; the latter is of key importance in order to make the best use of Europe's vast pool of talent.

3. **Adopt a definition of AI that captures what distinguishes AI approaches from other kinds of advanced computation: they exhibit key aspects of behaviour considered as intelligent in humans.** With a non-standard definition of AI, there is a risk that support as well as regulation are misaligned with what is commonly understood to constitute AI technology.

4. **Focus "AI made in Europe" on "AI for Good" and "AI for All"**; take global leadership, together with like-minded partners, in supporting publically funded, large-scale AI research and innovation that can compete at the level of large US and Chinese companies, while focusing on areas specifically relevant for societies.

5. **Establish a clear strategy for coordinating and structuring an AI-based innovation ecosystem across Europe.** Change existing policy instruments and strategies to take into account the significant role of entrepreneurs and private capital in the modern, AI-driven innovation economy.

6. **Establish policies to increase uptake of AI and investment in AI-driven product and market development** among the engines of the European economy.

7. **Invest in promoting broader awareness of AI in society**, and specifically of how AI technologies affect society and citizens; this is critical for the responsible use of AI and forms the basis for constructive engagement based on realistic expectations and adequate perception of risks.

8. **Build upon investments and tangible results of Horizon2020 programme in Responsible Research and Innovation (RRI)** to ensure that research and innovation in the field of AI achieve socio-economic benefits in Europe and strengthen democratic institutions, rule of law and human rights.

9. **Expand lessons learned in the areas of Privacy and Safety by Design in the last two decades and apply them to Ethics by Design for AI** by means of developing standards, metrics, legislation and institutional mechanisms for auditing, monitoring, inspection and certification.

10. **Create the proposed lighthouse centre in a way that effectively achieves critical mass, synergy, and cohesion across the European AI ecosystem** without permanently dislocating talent from where it is needed most. Make sure this is focussed on excellence and a site selection process grounded and transparently managed on the basis of politically neutral, externally validated criteria. Ensure this provides much-needed, large-scale data and computing infrastructure.

Introduction

The EU white paper on AI offers an initial framework for policymaking to foster the uptake of AI technologies in Europe and at the same time address the risks posed by the use of such technology in our society. **This framework is well aligned with the key elements of the vision developed and promoted since 2018 by CLAIRE.** CLAIRE is Europe's broadest AI community and the largest AI research network in the world, with more than 3000 individual supporters. CLAIRE's main purpose is to strengthen European excellence in AI research and innovation.

We believe that this white paper provides the appropriate framework to foster CLAIRE's vision of AI excellence in Europe, **including the establishment of a "lighthouse centre" – a vibrant hub for large-scale research and cooperation on AI in Europe** (similar in ambition and global recognition to CERN), a symbol for European ambition and achievement in this area, a global magnet for talent, and the centrepiece of an AI ecosystem that spans all of Europe and all areas of AI. **This lighthouse centre and the pan-European network served by it will play a key role in achieving critical mass, synergy, and cohesion across the European AI ecosystem – which is crucial for protecting and promoting European economic and societal interests globally**.

The approach outlined in the whitepaper is based on two pillars: an "ecosystem of excellence" and an "ecosystem of trust". While the first one establishes the actions to be taken to support the development and uptake of AI technology across the EU, the second one establishes necessary steps towards the creation of a regulatory framework for addressing potential risks and situations arising from the adoption of AI in our society.

Unfortunately, the white paper provides far less of a framework for supporting excellence in AI than for regulating AI; yet, **Europe cannot be a leader in AI regulation without being a leader in AI**; and **Europe cannot be a leader in AI applications or innovations without being a leader in AI research, including foundational research**. This implies that the EU must invest swiftly and substantially into Europe's "ecosystem of excellence", and that concrete plans for this (as discussed in more detail later in this document) need to be developed and put into action over the next few months.

To ensure that European AI systems, products and services can both comply with European norms and values and compete and take global markets, we need the capability to build them here. It is also the case that much of the basic research required to achieve this still needs to take place. This necessitates a substantial push by the very best AI researchers. The outcome of this research and the innovation enabled by it will give Europe an important competitive edge globally.

In addition to policy measures maintaining strong academic research in AI, they should be based on the best possible model of how AI-driven innovation largely happens. Innovation as a "value chain, starting in research and innovation", where the objective of policy measures is to "create the right incentives to accelerate the adoption of solutions based on AI" (White Paper, p. 3) does not reflect the current paradigm shift in innovation, a shift that is strengthened by AI.[2] In this game, Europe is not in the lead. Europe creates fewer new businesses destined for growth and

---

[2] This shift is recognised by the Commission's decision to establish the European Innovation Council, whose mandate will be to support disruptive innovation.

has few innovation ecosystems of strength and coherence. Europe has many small companies and startups, but – in contrast to the success of the typical European medium-sized and family-owned companies – very few of these scale up. It is therefore important to develop policy instruments that address this situation.

In the last 20 years, Europe's large strategic investments supporting R&D in robotics and their applications have leveraged Europe's top position in the world on robot systems incorporating AI. Work in this area will play a decisive role in addressing the economic, environmental and human challenges we are facing in areas such as healthcare, home robotics, autonomous cars, package delivery using unmanned vehicles in cities and many others. Thus, It is important to **establish a strong link between the areas of AI, Robotics and IoT,** since the potential of AI embodied in applications for the physical world is enormous, and Europe is well positioned to play a decisive role in that respect. Although we support the recognition of the role of hardware for AI "for the next data wave" (p. 4), we miss an explicit reference to the European strength in advanced sensors (laser sensors, high-precision video and audio sensors, GMR, AMR and TMR sensors), in cyber-physical, edge and near-field devices, which are as important for AI as the mentioned "low-power electronics", and in 5G communication.

In the context of the COVID-19 crisis, the EU has acted swiftly and with determination. We note that the broad adoption of AI will be much accelerated, world-wide, as a result of the major disruption caused by the COVID-19 situation. AI will be of key importance to the economic recovery and restructuring, and the key to achieving increased resilience in future crises. We expect that the COVID-19 crises will accelerate the broad adoption and roll-out of AI in industry and society. This means the risk for Europe to fall behind or become dependent on problematic AI systems has increased substantially, and so has the urgency to support AI research and innovation. **To retain, attract and regain key talent, and to achieve meaningful progress, substantive funding is needed, with streamlined allocation mechanisms focussed on those researchers and institutions with a track record of excellence, and on those with demonstrated potential for excellence.** The latter is of crucial importance, in order to prevent a situation where most resources are allocated to a small set of outstanding researchers, and to ensure the best possible development and utilisation of the enormous talent pool that exists across Europe. This includes not only supporting young researchers, but also initiatives designed to help less established research environments thrive at all scales.

We note that research, including **basic research and curiosity-driven exploration, are essential for the EU to maintain (and, ideally, solidify) its position as a major player in AI.** Europe will need all the excellence in foundational research it can access and build, in order to protect and advance key European economic and societal interests. In addition to large-scale research investments, we also recommend establishing a platform that enables a large pool of light-weight, smaller investments, in order to nurture early-stage, creative ideas, some of which may eventually evolve into large-scale projects in a bottom-up fashion.

The white paper also uses a **problematic and non-standard definition of AI**. While there is no universally accepted definition of what constitutes AI technology, defining AI as "a collection of technologies that combine data, algorithms and computing power" (p. 2) is problematic – especially, since it is later stated as forming the basis of any regulatory framework to be created (p. 16). While AI technologies make (to varying degrees) use of data, algorithms and computing power, so do many other types of information systems. Furthermore, while all AI systems rely on

sophisticated algorithms, some require large amounts of data, others large amounts of computation, and yet others both.

What distinguishes AI approaches from other kinds of computation is that they exhibit key aspects of behaviour considered as intelligent in humans, and thus enable fundamentally new levels of automation and delegation. The definition given in the white paper misses this point, and is also neither aligned with that given by the European Commission's High-Level Expert Group on AI, nor with that of the Association for the Advancement of AI (AAAI, the premier global scientific society for AI), which defines AI as "the scientific understanding of the mechanisms underlying thought and intelligent behaviour and their embodiment in machines".

AI thus encompasses algorithms and systems that can replicate, support or surpass human perceptual, linguistic and reasoning processes; learn, draw conclusions and make predictions based on large or small quantities of data; replicate or enhance human perception; support humans in diagnosis, planning, scheduling, resource allocation and decision making; and cooperate physically and intellectually with humans and other AI systems.

The definition used in the white paper does not reflect these diverse roles and capabilities of AI systems. It also does not consider that AI systems can be dynamic, non-deterministic, autonomous, adaptive and interactive. As a result, **it is unclear what needs to be supported and regulated, and there is a risk that support as well as regulation is misaligned with what is commonly understood to constitute AI technology.**

We appreciate that the White Paper stresses "the need to build bridges between disciplines […] such as deep learning and symbolic approaches" (p. 5), but it is an oversimplification – often found in the popular press – that symbolic approaches are "rules created through human intervention". In addition to production rules, symbolic methods include logical, spatial and temporal reasoning, automated theorem proving, semantic networks and knowledge graphs, constraint solvers, as well as qualitative models.

Europe has the opportunity to establish a unique approach to AI, by combining the key notion of human-centric AI with those of "AI for Good" and "AI for All" – concepts that are rapidly gaining attention world-wide. It is very good to see this clearly reflected as the basis for the ecosystems of excellence and trust sketched out by the white paper.

It is the right strategy for the EU and its partners to focus on the development and use of AI for the good of our citizens, our society and our planet. We should aim at creating intelligent machines that implement fundamental and shared values, respect and amplify human abilities and support the shaping of a better society. We should maximally leverage AI for achieving the UN Sustainable Development Goals – **"AI made in Europe" should be "AI for Good"**.

It is also important to embrace the diversity of the different regions and cultures in Europe, making sure that the AI framework benefits all of Europe and leverages the talent and resources our diverse regions and societies have to offer. The European approach to AI should foster the accessibility of knowledge and broadly deployed technology by everyone, across different generations, with or without specialised education, by lowering the barrier to entry for the effective, safe and beneficial use of AI – **"AI made in Europe" should be "AI for All".**

This also implies that vigorous EU antitrust and fiscal laws, and the compliance thereof, for powerful market parties need to be pursued. This is essential to address the inequalities and injustices that are typically associated with dominance of powerful parties in markets and ecosystems of AI and data.

Overall, we see an excellent opportunity for Europe to establish itself as a leader on a globally impactful, public push for "AI for Good" and "AI for All", which complements the commercial push of for-profit AI and ensures that the public has a stake in this key enabling set of technologies. AI is too important to be controlled by industry alone. This is in analogy to the public effort on sequencing the human genome in the late 1990 and of even broader importance, considering the anticipated impact of AI on all areas of human endeavour. **Europe needs to be able to support large-scale AI research and innovation accessible to the public and accountable to the public, research and innovation that can compete at the level of large companies, while focusing on areas specifically relevant for our societies.** This also implies that Europe should adopt a cautious approach to military applications of AI (especially offensive capabilities), should hold back on the development of Lethal Autonomous Weapons Systems, while taking part in UN discussions in Geneva about AI and International Humanitarian Law and in general insist on peaceful applications of AI.

An ecosystem of excellence

With a coordinated world-class research community collaborating with leading universities and companies around the world, Europe is well positioned to play a leading role in AI research and application. Without research excellence there can be no major AI innovations or trust in AI technologies developed in Europe. The Commission has rightly identified a need for focus on investment in and deployment of AI to ensure its benefits across the board continue to grow, in particular in the context of the future economic recovery. Below are our recommendations to take the next step and ensure the vision of an ecosystem of AI excellence in Europe.

Invest in promoting a broader awareness of AI in society, and specifically of how AI technologies affect society and citizens; this is critical for the responsible use of AI and forms the basis for constructive engagement based on realistic expectations and perception of risks. It also strengthens the autonomy of people, in line with the first principle of the European Ethics Guidelines for Trustworthy AI, the respect for human autonomy.

Invest in both up-/reskilling and in basic education related to AI. Education is fundamental to reach excellence. Europe already has a good educational system that can be further improved. First, there is a need to significantly increase the volume of **broad AI educational programmes with a focus on technology** (at all levels including BSc, MSc, PhD, and postdoctoral). Second, develop **specific AI educational programs with a focus on dissemination in other sciences and society as a whole** (at all levels including BSc, MSc, PhD, and postdoctoral). Third, make sure that **primary and secondary education** provides the necessary theoretical and practical foundations to allow everyone to become active and engaged citizens in the modern society, where AI is a natural part.

Both mission-oriented and fundamental research in all aspects of AI must be secured in order to promote AI that is trustworthy and to address relevant scientific, ethical,

sociocultural and industrial challenges. The overarching goal is to ensure world-class research that contributes to Europe's sustainability, growth and competitiveness, and inclusion, leading to individual and societal well-being. This necessitates a European research community that can unite through strong collaboration, and that can join forces with industry and society at large to build on European research strengths and enhance Europe's well-being.

Provide dedicated, significant and long-term research funding for both fundamental and mission-oriented research on AI to maintain competitiveness of European research and companies, and to address relevant societal and industrial challenges. The funding should be independent of, and parallel to, any funding programme on ICT. Significant investment in basic funding is required to achieve adequate research independence, which is necessary for fundamental excellent research in AI.

A European AI Lighthouse. It should be "**the place to be**" when it comes to **AI research and innovation in Europe**. A place where people can meet for a period of time to work with other leading researchers and experts from all over the world on the most exciting and important topics, technologies and applications of AI. Through sabbaticals and other temporary scientific positions, the Hub will not drain talent from labs around Europe. Rather, it will act as the **beating heart of European AI**, a **place where knowledge is exchanged, fused and amplified** by the visiting researchers **and then spread out** again to the labs in the network by the returning researchers, thereby strengthening the development of excellent AI research across all of Europe.

The ambition should be to create a European institution with the global impact and brand recognition of CERN and ESA. Towards this end, it is crucial to involve an international group of renowned AI experts, including experts from outside of Europe, in the design and realisation of the concept. The thorny issue of site selection should be based on objective criteria developed by one such expert group and a competitive, transparent process based on these criteria, with the goal of best achieving the intended impact of the new lighthouse centre. Criteria that are important in this context likely include: the ability of the site to accommodate a world-class facility; the strength of the local AI ecosystem; ease of access (proximity to major transportation hubs); international appeal of the location and attractiveness to the world-class staff that will work there; support by national, regional local governments and companies.

Create incentives and support for interdisciplinary and multi-stakeholder research, for example through large-scale challenge-driven research missions. While striving for technological leadership, also ensure the involvement of social scientists, humanities scholars, as well as other interdisciplinary collaborations in European AI research projects. As AI is having more and more impact on society at large, there is a corresponding increase in the need to study these aspects from humanistic and social scientific perspectives. This is also an area where Europe has a well established track record. The WASP-HS programme in Sweden and the Humboldt professorship programme in Germany are good examples. Foster the involvement of civil society in AI research programme design and in projects (e.g., through citizen science) where appropriate.

Simplify and streamline the structure of research funding instruments. Reduce the overhead in efforts and time to decisions of applying for grants, for example through early reviews from abstract-sized submissions; improve the quality and consistency of proposal reviews, and increase the predictability of funding instruments and their timing to support long-term planning. Since top

researchers produce excellent results with high consistency, a substantial part of the funding should be allocated based on track record (e.g., similar to the Reinhart Koselleck funding in Germany for excellent researchers, which is based on 5-page proposals for up to 1.25 M€) rather than the need for extensive research proposals. Of course, substantial funding opportunities also need to be provided for talented junior researchers without a long-standing track record.

Adopt current models of innovation. "Technology transfer" from academia to large companies for commercialisation is just one of several innovation processes in the innovation economy. First, in Europe, medium-sized, often very specialised companies, are often global export champions and drivers of innovation. Together with large companies, for which they serve as suppliers, and a network of start-up companies, they form a unique European innovation system in the private sector. Second, entrepreneurs, startups and investors also play an important role. Today, innovation stems largely from research laboratories, digital platforms, medium-sized companies, and startups and is largely an emergent effect of non-linear interactions between heterogeneous actors in an ecosystem where entrepreneurs and investors play just as an important role as industry, academia and governments. Innovation is also largely driven by medium-sized companies, startups, and private capital. Digital Platforms and AI-based disruptive business models play a key role in the innovation portfolio. A mismatch between the economy and the policy instruments may not only be useless, it may do more damage than good. Given the comparatively lower success rate of European AI-based startups but the resilience and dynamics of the medium-sized companies, we believe this should be more strongly reflected in the Commission's AI strategy.

Develop new policy instruments or adapt existing ones that reflect the realities in the modern innovation economy. AI is a key technology area in the new innovation economy. At the same time, the new innovation economy is key to AI's strength in Europe – including for applied and foundational research in academia. The future of AI in Europe is therefore dependent on Europe's ability to develop its innovation economy. Today, that means developing the innovation ecosystems to foster data- and AI-driven startups and grow them to scale[3], which again means bringing the entrepreneurial and private capital segments into the design of policy instruments and strategies. Examples include deepening corporate-startup collaboration, increasing the use of entrepreneurial student projects in collaboration with industry, developing publicly available makerspaces, opening funding instruments to wider audiences, establish common labs and arenas for ecosystem actors, including universities, and establishing a clear strategy for coordinating and structuring a cohesive European AI ecosystem. **The interaction between fundamental research and other functions in the innovation ecosystem needs to be substantially increased, and time from research to market needs to be shortened. European AI centres should be established, with the mission of building and growing the European AI innovation ecosystem.**

Responsible innovation and ethics by design approach. Europe has built strengths in bringing ethics and moral values effectively and demonstrably to bear upon engineering and technology. Safety by Design and Privacy by Design have been pioneered in Europe. The future will not only need design for fairness, accountability, equity, and non-discrimination, but also needs to develop

---

[3] European companies seldom grow to scale. Far too many remain two-person, three-person, or quite often just one-person companies. Their innovative ideas remain the exclusive domain of local economies, sometimes confined to a single European Union member state, sometimes even to a single region within them.

methodological approaches and tools to support engineering design for values. Ethics need to be present when and where the world is shaped.

Build upon investments and tangible results of Horizon2020 programme in Responsible Research and Innovation (RRI) to ensure that research and innovation in the field of AI achieve socio-economic benefits in Europe and strengthen democratic institutions, rule of law, and human rights. It has been shown that constraints can actually be beneficial for the economy and stimulate innovation.[4] In order to have a significant impact on applications and actively stimulate responsible innovations, funding should be targeted towards existing scientific strengths, novel research opportunities, and key European interests.

Adopt AI technologies across all industries and sectors in order to benefit from competitive gains. In particular, policies should be considered to accelerate uptake of AI and investments in AI-driven products and services among the engines of the European economy. There are indications that important European industries and areas of technology, including energy, the automotive industry, the Internet of Things, and robotics, have not yet embraced the shift to AI. For instance, there are few European AI startups connected with those sectors.[5]

An ecosystem of trust

The Commission is right to emphasise and highlight the importance of Trust for the future development of AI in Europe. We are happy to see that the commission moves beyond the language of "trustworthy AI", an expression that suggests that inherent properties of the technology may provide sufficient grounds to establish relationships of trust. Trust is correctly situated in the white paper in the context of *ecosystems*, since it pertains to a fabric of social, institutional and legal preconditions that are necessary for high quality and low transaction cost relationships between human beings. The Commission construes Trust not so much as a property of the technology, we believe correctly, but in a broader and more original sense, as a property of human relations and a quality of expectations of persons.

The expression "Trustworthy AI" recently gained currency and enjoys popularity, but it may not do full justice to what is required to prepare European society for the Age of AI; we suggest the idea of human centred-ness equally applies. Humans may *rely* on technology and may have (or lack) *confidence* in technology, which means that they do (or don't) believe that the technology will work in accordance with known specifications and requirements.

Trust is however more than mere reliance or belief in reliable performance. It implies that the trusting person believes and expects that other persons will act from a moral point of view, not in the least exactly because the trustee believes that she is trusted and depended upon. The person that is being trusted typically values that he or she is deemed worthy of trust and is more likely to behave accordingly. Trust is thus a fairly delicate and complex moral feature of human relationships.

It is of the utmost importance that an ecosystem of Trust accommodates this rich human phenomenon. Legal, institutional and social arrangements and mechanisms of attributing and apportioning of responsibility are therefore to be designed with a focus on human

---

[4] HBR Nov 2019; https://hbr.org/2019/11/why-constraints-are-good-for-innovation
[5] "Artificial Intelligence A strategy for European startups", Roland Berger, Asgard – Human Venture Capital, 2019

beings and aim to enhance moral lives and moral resources of individual citizens as moral agents. This aspect of human-centred AI foregrounds human agency, since only human beings can strictly speaking be subjects of moral agency, moral responsibility and moral personhood. Organisations and institutions, systems and technology can be trusted in a derivative sense, but that should not abstract from the fact that trust is a moral notion and therefore essentially human-centred.

One of the prime virtues of ecosystems of trust is that they are designed to make unambiguous allocations of responsibility to human agents possible. The design and development of an ecosystem of trust should not repeat the mistakes that have been made in the financial world where intransparent legal, corporate governance and fiscal constructions have contributed greatly to obscuring and concealing human agency and limiting moral responsibility. This has undermined trust in the financial and corporate world.

This connects to Europe's strong and perhaps world-leading position in the field of formal verification – i.e., formal proof assistants, theorem provers and satisfiability-based systems, model checkers, and related methods. Formal verification is a critical technology for ensuring with mathematical certainty correct functioning of software, hardware and other complicated designs. Apart from ensuring bug-free mission-critical programs and operating systems, formal verification research has recently targeted complex blackbox AI systems such as neural networks, designing methods that ensure that such blackbox systems satisfy required safety properties.

The long-term social and economic advantages of high trust are obvious. Europe has perhaps the strongest legal and institutional framework internationally, with its GDPR, E-Privacy regulation, Charter of Fundamental Rights, and the European Convention on human rights and related laws. Adding to these, by fostering ecosystems of Trust in the Age of AI, is the greatest service that can be done to the future development and uptake of AI, and to the economic prosperity enabled by that.

Safety and liability implications of AI, IoT and Robotics

The commission rightly focuses on the need for conceptual clarification and legal development in the context of the safety risks of AI. The white paper provides a good and comprehensive overview of the safety risks and identifies the conceptual gaps in the application of the extant EU law on product safety, consumer protection and liability. The white paper shows that the Commission is dedicated to applying the highest standards of safety and risk mitigation to AI, without unnecessarily creating obstacles to high-quality research.

We think that the direction that the Commission has taken by distinguishing different types of applications in different sectors of society is fruitful. **The idea that there are high-risk applications and low-risk applications is sensible.** Fairly straightforward routing tasks in logistics should obviously meet different safety and liability standards than robotic surgery applications. The listing in the white paper of safety risks in a broad range of applications and the requirement of prior conformity assessment for high-risk applications is a viable approach. However, **we need to bear in mind that the application of thinking about liability, causation, control, and contributory fault in the context of AI is not a simple exercise and will require considerable legal and philosophical research.** It is important in that context to create alignment in thinking about responsibility and liability between causation, human agency, control,

and cost of contributory faults; how these should be conceived in the case of AI and ML is not a simple matter.

The general approach thus is not to provide a framework with exclusively a long list of general principles, but instead differentiate socially, institutionally and technologically. Over the centuries we have made societies safe not by introducing lists of general principles, but by introducing very specific, detailed and effective institutions (food and drug administration, aviation and maritime safety boards, accident investigation boards) and legal frameworks that pertain to them.

Elaborate schemes and mechanisms of monitoring, inspection, certification, quality control and auditing have typically been tied to sectors and types of applications. We now can board a plane, buy a sandwich, take a medicine, drive our car and cross a bridge without much thought, since we know there is a solid and sophisticated institutional, legal and social framework that pertains to the said activity, and this framework is complemented by human oversight, correction and social learning. We have now entered a new world of advanced AI that is likely to penetrate every nook and cranny of society. We still have to provide a system of liability risk and safety management at the collective and societal level that makes our societies safe for the large-scale introduction of AI. **We have to design an elaborate system that is the functional equivalent of our best practices and exemplary paradigms in safety management and certification in the food and drugs, civil engineering, transport and mobility, and the consumer protection domains, and we have to do it fast.**