Response to the European Commission's Consultation on Artificial Intelligence: A European approach to excellence and trust

Response to the European Commission's Consultation on Artificial Intelligence: A European approach to excellence and trust

This document is a response to the European Commission's Consultation on Artificial Intelligence, from Loughborough University systems engineering researchers Dr Melanie King and Paul Timms, written as part of the TECHNGI Academic Research Project (UKRI Project Ref: ES/S010416/1). TECHNGI (Technology Driven Next Generation Insurance) is a cross-disciplinary research project investigating the opportunities and challenges for the UK insurance industry arising from the application of new AI technologies, including machine learning, distributed ledger, automated processing, and the explosion of available data[a].

We provide both general comments on the white paper [1], and address more specific responses to the three main sections in the survey:

- **Section 1 – An ecosystem of excellence** refers to the specific actions, proposed in chapter 4, for the building of an ecosystem of excellence that can support the development and uptake of AI across the EU economy and public administration.
- **Section 2 – An ecosystem of trust** refers to a series of options for a regulatory framework for AI, set up in chapter 5.
- **Section 3 – Safety and liability implications of AI, IoT and robotics** refers specifically to the companion report on the safety and liability aspects of AI [2].

Our response (this paper) to the European Commission is structured as follows:

[a] For more information about the TECHNGI research project, please visit http://www.techngi.uk.

Background to the consultation

The European approach for Artificial Intelligence (AI) aims to promote Europe's innovation capacity in AI while supporting the development and uptake of ethical and trustworthy AI across the European Union (EU). According to this approach, AI should work for people and be a force for good in society. The European Commission's (EC) Digital Strategy (2019 – 2024) [3] has prioritised action in three key areas [4]: Excellence and trust in artificial intelligence, European data strategy, and the European industrial strategy.

As part of the Digital Strategy the EC proposes:

- that new legislation on AI should be adapted to the risks, it should be effective but not limit innovation;
- to require high-risk AI systems to be transparent, traceable and under human control;
- that authorities must be able to check AI systems as they check cosmetics, cars or toys;
- to ensure unbiased data sets;
- to launch an EU-wide debate on the use of remote biometric identification (e.g. facial recognition).

EC' Guidelines on Trustworthy and Transparent AI

In April 2019, the European Commission's (EC) High-Level Expert Group on AI presented ethics guidelines for trustworthy artificial intelligence [5]. This followed the publication of the guidelines' first draft in December 2018 on which more than 500 comments were received through the first round of open consultation [6].

Based on fundamental rights and ethical principles, the Guidelines list seven key requirements that AI systems should meet in order to be trustworthy [5]. These requirements are applicable to different stakeholders within AI systems' life cycle; these include developers, deployers and end-users, as well as the broader society.

Below is a list of (non-exhaustive) requirements set out in the guidelines, which includes systemic, individual and societal aspects:

Human agency & oversight	Including fundamental rights, human agency and human oversight
Technical robustness & safety	Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility
Privacy & data governance	Including respect for privacy, quality and integrity of data, and access to data
Transparency	Including traceability, explainability and communication
Diversity, non-discrimination and fairness	Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation
Societal & environmental wellbeing	Including sustainability and environmental friendliness, social impact, society and democracy
Accountability	Including auditability, minimisation and reporting of negative impact, trade-offs and redress

EC White Paper on Artificial Intelligence (Feb 2020)

The focus of the current public consultation (via an online survey) is on the *"White Paper on Artificial Intelligence – A European Approach"* [1], published in February 2020, which aims to foster a European ecosystem of excellence and trust in AI and a report on the safety and liability aspects of AI. The White Paper proposes:

- Measures that will streamline research, foster collaboration between Member States and increase investment into AI development and deployment;
- Policy options for a future EU regulatory framework that would determine the types of legal requirements that would apply to relevant actors, with a particular focus on high-risk applications.

The White Paper and the survey also draws upon two complementary EC publications: the *"Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics"* [2] *and "A European strategy for data"* [7]. The consultation is open from February to June 2020 with the aim of implementing new policy and regulatory requirements towards the end of 2020.

This document provides a response from a systems perspective to the EC's online survey questions, available online at:

https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12270-White-Paper-on-Artificial-Intelligence-a-European-Approach/public-consultation

General comments on the EC's approach within the white paper

THE WHITE PAPER

There are two thematic strands that provide structure strategy, these are **excellence** and **trust**. Within the context of the white paper [1] under consultation, these two themes are used to separate the issues of *value creation from data* ('excellence') and the need for a *human-centric approach to data use* ('trustworthiness'). This separation is presented in the form of two ecosystems [1]:

an ***ecosystem of excellence*** – that "supports the development and uptake of AI across the EU economy and public administration"; and

an ***ecosystem of trust*** – that addresses "the socioeconomic aspects in parallel with an increase in investment in research, innovation and AI-capacity across the EU".

We believe: The ecosystems of excellence and trust are not separate

The terms ecosystem of excellence and ecosystem of trust are used in the white paper to divide the *value drivers of AI* from *the activities that will inhibit such value creation*. However, in doing so, the report implies two assertions, that we believe are incorrect:

- That value creation is always excellent (i.e. it is always a positive measure for all stakeholders).
- That trustworthiness can only be achieved through regulation[b].

The idea of ecosystems of *excellence* and ecosystems of *trust* are not in themselves poor notions, but they are more usefully presented and understood differently than in the White Paper. Trust, in the context of the High Level Expert Group's definition of an AI system [8], can be thought of as a property (or capability) of the technology-enabled system. Excellence, by contrast, is a measure of how well a specified quality is achieved. Consequentially:

- An ecosystem of trust is one part of an ecosystem of excellence
    - The continued presentation of excellence and trust in their current form (as separate ecosystems) will, we fear, lead to disaggregation between the policies and programmes of work focussing on value creation, and those lines of work focussing on the regulation of technology.
    - As we further elaborate on page 19, an ecosystem of regulation is one part of an ecosystem of trust.
- The 'ecosystem of excellence', as it is currently presented in the white paper, is actually an 'ecosystem of data value creation'.
    - The adoption of AI is an issue driven by ethical decisions and moral trade-offs between various values. It is well-acknowledged in AI ethics-related literature[c] that, while the many principles of a 'good' ethical system are known (such as the Commission's seven key requirements [5]),

[b] See page 18 for a further rationale behind this statement.
[c] In particular, we draw the Commission's attention to Martin et al's [30] paper on the Business and Ethical Implications of Technology.

it is the choices made in the trade-off between principles, caused by a wide array of moral dilemmas, that will characterise whether the ecosystem is considered 'excellent' by its stakeholders. It is unlikely that the 'average member of the public' will trust a system whose only demonstrable values are those that benefit business. In its current form, the inference is that the European Commission prioritises the benefits of AI for business, which in some circumstances might be at odds with the benefits and needs of society (see comments below).

While it is understandable that the authors of the whitepaper may wish to separate the issues of value creation and regulation in order to manage the complexities of AI adoption, the separation of the themes of 'excellence' and 'trust' go against well-established definitions of these terms. This may lead to greater confusion in an area of technology adoption already suffering from a definition crisis.

We believe: The report focuses too heavily on economic benefits for business, to the detriment of AI as an enabler for society as a whole

The political guidelines of Commission President Ursula von der Leyen, amongst a range of policy objectives, call for *"an economy that works for people"* and *"a Europe fit for the digital age"* [9]. in Leyen's political guidelines for *"An economy that works for people"*, reconciliation is sought between *"the social and the market in today's modern economy"* [9]. In chapters 2 ("Capitalising on strengths in industrial and professional markets") & 3 ("Seizing the opportunities ahead: the next data wave") of the white paper, the EC's primary driver for AI adoption appears to be to increase its economic standing within a global context. This position is reinforced through statements such as:

- *"Harnessing the capacity of the EU to invest in next generation technologies and infrastructures, as well as in digital competences like data literacy, will increase Europe's technological sovereignty in key enabling technologies and infrastructures for the data economy"*.
- *"Although Europe currently is in a weaker position in consumer applications and on online platforms, which results in a competitive disadvantage in data access, […] Each new wave of data brings opportunities for Europe to position itself in the data-agile economy and to become a world leader".*

The EU's driver is to increase its global economic standing and was founded on the idea of a common market. Therefore, EC's strategy - to position itself to become a world leader in the data-agile economy, addressing its current 'weak position' in consumer applications and digital platforms - is understandable. Their political and economic aim is to address the current "competitive disadvantage in data access, major shifts in the value and re-use of data across sectors" and compete with the USA's BigTech firms and China's tech dominance. However, as noted within the paper, "the impact of AI systems should be considered […] from the perspective of **society as a whole**" [1], and chapters 2 & 3 of the white paper, which frame the benefits for the remainder of the paper, do not really do this.

Furthermore, in chapter 5 ('ecosystem of trust'), while we acknowledge that the proposed regulatory framework will afford various protections to society, we must again emphasise that, the discussion is again focused on the limitation of certain profit-making activities, rather than on the enablement of positive AI uses. Trust is achieved through positive action, and we believe the EC should be regulating, not only to limit harmful instances of AI, but to encourage adoption of beneficial uses of AI (i.e. the scope is wider than product safety legislation).

For example, on page 2 of the white paper, reference is made to the Green Deal and the World Economic Forum's sustainability goals. On page 10, it is noted that AI can make products and processes safer. The white paper would benefit from articulating these benefits (and others) more specifically.

However, societal benefits are wider than just sustainability. We draw the Commission's attention to Floridi et al's paper on an Ethical Framework for a Good AI Society [10]. In this report, they draw attention to four key societal opportunities of AI; (1) Enabling human self-realisation, (2) Enhancing human agency, (3) Increasing societal capabilities and (4) Cultivating societal cohesion. The European Union could play a valuable part in the betterment of society, through the chosen policies and regulatory decisions made through this work, encompassing these four factors. In its current form, the EC's white paper does not address these opportunities.

Response to Section 1 – An ecosystem of excellence

Question 1: Actions of the European Commission

THE WHITE PAPER – Section 4 'An ecosystem of excellence'

To build an ecosystem of excellence that can support the development and uptake of AI across the EU economy, the white paper proposes a series of actions. There is a total of six actions falling within eight strategy considerations:

A: Working with member states

Action 1: The Commission, taking into account the results of the public consultation on the White Paper, will propose to the Member States a revision of the Coordinated Plan to be adopted by end 2020.

B: Focusing the efforts of the research and innovation community

Action 2: The Commission will facilitate the creation of excellence and testing centres that can combine European, national and private investments, possibly including a new legal instrument. The Commission has proposed an ambitious and dedicated amount to support world reference testing centres in Europe under the Digital Europe Programme and complemented where appropriate by research and innovation actions of Horizon Europe as part of the Multiannual Financial Framework for 2021 to 2027.

C: Skills

Action 3: Establish and support through the advanced skills pillar of the Digital Europe Programme networks of leading universities and higher education institutes to attract the best professors.

D: Focus on SME's

Action 4: The Commission will work with Member States to ensure that at least one digital innovation hub per Member State has a high degree of specialisation on AI. Digital Innovation Hubs can be supported under the Digital Europe Programme.

E: Partnership with the private sector

Action 5: In the context of Horizon Europe, the Commission will set up a new public private partnership in AI, data and robotics to combine efforts, ensure coordination of research and innovation in AI, collaborate with other public-private partnerships in Horizon Europe and work together with the testing facilities and the Digital Innovation Hubs mentioned above.

F: Promoting the adoption of AI by the public sector

Action 6: The Commission will initiate open and transparent sector dialogues giving priority to healthcare, rural administrations and public service operators in order to present an action plan to facilitate development, experimentation and adoption. The sector dialogues will be used to prepare a specific 'Adopt AI programme' that will support public procurement of AI systems, and help to transform public procurement processes themselves.

G: Securing access to data and computing infrastructures

No actions emphasised.

H: International aspects

No actions emphasised.

Question 1. In your opinion, how important are the six actions proposed in section 4 of the White Paper on AI?						
	Not important at all	Not important	Neutral	Important	Very important	No opinion
Working with Member states					X	
Focussing the efforts of the research and innovation community					X	
Skills					X	
Focus on SMEs					X	
Partnership with the private sector					X	
Promoting the adoption of AI by the public sector					X	

Our response: All of these actions are very important

We think all of these actions are very important. The critical concern is not the areas in which actions are taken, but rather how this action is pursued. For example it is important to ensure that the efforts on the research and innovation community include research on the effective application of AI in its business or public policy context, it is not possible to assess the excellence and value creation of AI technologies in isolation from their context.

Question 1a. Are there any other actions that should be considered?

Our response: There is a need to engage more broadly with the public

Arguably, it is the responsibility of the European Commission to ensure engagement with all stakeholder groups that will be affected by the increased adoption of AI technologies. Actions 1 to 6 represent a reasonable cross-section of stakeholder interests, however, there appears to be a clear omission from this list of actions, in the form of engagement with consumer advocacy groups. This omission is also present within the actions of the "coordinated plan on AI" (section 4A).

While it is acknowledged that the consultation itself enables all European citizens to "provide their opinion on the white paper and contribute to the European approach to AI", this is not a rigorous approach to ensuring that future use of AI is within the publicly-acceptable boundaries of AI use. While it is acknowledged that consumer expectations will be partially represented through consideration in the "ecosystem of trust", these two ecosystems cannot be viewed and managed as separate entities (as we have previously argued). To assume that commercial or government entities with vested interests would accurately and comprehensively represent consumer interests in AI would be foolhardy.

[NB: While we have no particular commentary to provide regarding these categories, it seems an odd omission that the section headings for' G: Securing access to data and computing infrastructure' and 'H: International aspects' do not come with associated actions on the part of the EC.]

Question 2: Revising the coordinated plan on AI to be adopted by the end of 2020

THE WHITE PAPER – Section 4a 'Working with member states'

In section 4a of the white paper, the main activities of an updated "coordinated plan" (an expansion on Action 1) are proposed: (1) Strengthen excellence in research, (2) Establish world-reference testing facilities for AI, (3) Promote the uptake of AI by business and the public sector, (4) Increase the financing for start-ups innovating in AI, (5) Develop skills for AI and adapt existing training programmes, (6) Build up the European Data Space.

Question 2. In your opinion, how important is it in each of these areas to align policies and strengthen coordination as described in 4.a of the white paper?						
	Not important at all	Not important	Neutral	Important	Very important	No opinion
Strengthen excellence in research					X	
Establish world-reference testing facilities for AI					X	
Promote the uptake of AI by business and the public sector					X	
Increase the financing for start-ups innovating in AI					X	
Develop skills for AI and adapt existing training programmes					X	
Build up the European data space					X	

Our response: All of these areas are very important

We think all of these actions are very important. We would add, from the experience of our TECHNGI project working on the application of AI in insurance, that it is critically important to support co-operation between the private sector, the public sector and the research and innovation community in all these areas. Only when stakeholders of significantly different world views are brought together will we reach an acceptable balance of AI use that benefits all parties.

Question 2a. Are there any other areas that should be considered?

Our response: There is a need to align policies for the promotion of societal welfare and environmental benefits.

As identified above we feel that consumer, and more generally, public engagement is lacking within the EC's "ecosystem of excellence". In chapter 4a of the white paper, it is stated that "The Coordinated Plan could also address societal and environmental well-being as a key principle for AI", however the 6 proposed activities (which we summarise above) do not contribute to this aim. We feel that the coordinated plan ***should*** address societal and environmental wellbeing, rather than ***could***.

It should be emphasised that societal and environmental wellbeing is wider than the presented context of "climate change and environmental degradation". As we argue elsewhere within this paper, the benefits to society are somewhat lacking within the EC's white paper. We don't seek to provide a comprehensive list of benefits here, but draw the reader's attention to Adler & Seligman [11], and in particular their argument:

> *"rather than targeting GDP growth, national governments can provide the enabling conditions for wellbeing through better public services (e.g., health and education), urban planning that promotes relational leisure and diminishes commuting times, and a stronger social safety net. As others have before them (e.g., Okun, 1975), they acknowledge that there are trade-offs between different contributors to wellbeing, including freedom, opportunity, efficiency and equality. To best enable social wellbeing, each government needs to weigh up these trade-offs, depending on what constitutes value".*

The benefits of AI and data capabilities now available to corporate and public bodies are recognised in the EC's strategy for data [7], however these should be addressed across publications, for fear of siloed contextual awareness.

Our response: Inclusion of independent consumer representation bodies in all areas

While government intervention and regulation of AI use will, in itself, garner an increased level of trust in AI use (within industries), this is only possible if the 'general public' trust the motives and involvement of those in positions of enablement. This includes the public institutions of government and politics themselves. Once such mechanism to ensure trust would be to strengthen the representation of 'the general public' through mandatory engagement with (independent) consumer advocacy groups to ensure balanced representation 'around the table'.

Question 3: A united and strengthened research and innovation community

Question 3. In your opinion, how important are the three actions proposed in sections 4.b, 4.c and 4.e of the white paper?						
	Not important at all	**Not important**	**Neutral**	**Important**	**Very important**	**No opinion**
Support the establishment of a lighthouse research centre that is world class and able to attract the best minds				X		
Networking of existing AI research excellence centres					X	
Set-up of public-private partnership for industrial research			X			

Our response: Networking and collaboration is of particular importance

As discussed earlier in our response to Question 2, cooperation is required between research, public institutions and industry.

Question 3a. Are there any actions to strengthen the research and innovation community that should be given a priority?

Our response: Developing instruments to facilitate co-operation between the users of AI in both the private and public sectors and the research and innovation community

As discussed earlier in our response, it is essential to see AI in its application context; therefore the work of the research and innovation community must be pursued in close co-operation with users.

Question 4: Focusing of Small and Medium Enterprises (SMEs)

Question 4. In your opinion, how important are each of these tasks of the specialised Digital Innovation Hubs mentioned in 4.d of the white paper?						
	Not important at all	**Not important**	**Neutral**	**Important**	**Very important**	**No opinion**
Help to raise SME's awareness about potential benefits of AI			X			
Provide access to testing and reference facilities					X	
Promote knowledge transfer and support the development of AI expertise for SMEs			X			
Support partnerships between SMEs, larger enterprises and academia around AI projects					X	
Provide information about equity financing for AI startups						X

Our response: AI depends on access to good quality data, skilled engineers and scientists as well as access to computational capacity; in order to develop, train and rapidly test model development. This is a particular challenge for SMEs, requiring partnership between SMEs, larger enterprises and academia.

Our TECHNGI project is examining the crucial challenge of sharing and opening access to data to support AI applications. This requires co-ordination between SMEs, larger enterprises and academia. "Partnership" and "project" are rather too narrow a concept to be applied here. This can be better phrased more broadly as "co-ordination of effort" between SMEs, larger enterprises and academic around AI initiatives".

Question 4a. Are there tasks that you consider important for specialised Digital Innovation Hubs?

Our response: Promoting the development of accessible, useable data.

A key finding emerging from the TECHNGI project is that hubs play a critical role of supporting data to digital innovation (to quote a phrase from IBM highlighted at a Nov 2019 TECHNGI project conference in "There is not Artificial Intelligence AI without Information Architecture IA" [12].

Further observations on Section 1

While we have no particular commentary to provide relating to Actions 3, 5 or 6, it seems an odd omission that the actions for ***skills***, ***partnership with the private sector*** and ***promoting the adoption of AI by the public sector*** are not included in the consultation questions.

Comments on Section 2 - An ecosystem of trust

THE WHITE PAPER – Section 5 'An Ecosystem of trust: Regulatory framework for AI'

In section 5 of the white paper a series of options for a regulatory framework for AI are presented.

In the introduction of this section, it is put forward that: (1) AI brings both opportunities and risks, (2) Citizens fear […] the information asymmetries of algorithmic decision-making, (3) companies are concerned by legal uncertainty, (4) citizens worry that AI can have unintended effects or even be used for malicious purposes. Reference is made to previous EC publications on the EC's AI Strategy [11] (forerunner to the current EC whitepaper under consultation) and EC publications on guidelines for trustworthy AI [13], [14].

The remainder of this section puts forward the case for a "clear European regulatory framework [to] build trust among consumers and businesses in AI, and therefore speed up the uptake of the technology". The EC argues that such a framework:

- Should be consistent with other actions to promote Europe's innovation capacity and competitiveness in this field.
- Must ensure socially, environmentally and economically optimal outcomes and compliance with EU legislation, principles and values.
- Must leave room to cater for further developments in the evolution of AI capability.

Question 5: Options for a regulatory framework

Question 5. In your opinion, how important are the following concerns about AI?	Not important at all	Not important	Neutral	Important	Very important	No opinion
AI may endanger safety					X	
AI may breach fundamental rights (such as human dignity, privacy, data protection, freedom of expression, workers' rights etc.)					X	
The use of AI may lead to discriminatory outcomes					X	
AI may take actions for which the rationale cannot be explained					X	
AI may make it more difficult for persons having suffered harm to obtain compensation					X	
AI is not always accurate					X	

Our response: All of these areas are very important

All of these actions can be traced back to the principle of 'do no harm' (non-maleficence). As such, we argue that there should be no differentiation between these issues.

Question 5a. Do you have any other concerns about AI that are not mentioned above?

Our response: Concerns regarding AI are better understood when the 'owner' of that concern is defined

This section could benefit from greater emphasis on whom owns the concern, is it business, consumer, society, government, etc. We assume, from the context of the issues presented, that this section is about societal demographic risks. However, in the principle of good stakeholder management, all perspectives should be understood and considered. The relationship between stakeholder groups should also be understood, as some requirements and ethical principles will conflict. The risk of market concentration (see below) will reduce competition, which in turn may increase the cost of goods and services and impact on the available choice to consumers, causing a detrimental impact to price.

Our response: There are additional concerns about AI

As introduced in our response to question 5, there are many ways in which AI can cause harm. As such, AI needs to be deployed in a controlled and measured way. We applaud the EC's acknowledgement that AI capability is ever evolving and agree that, within an AI-application context (e.g. insurance), a comprehensive set of risks are still evolving due to the infancy of application. We hope that that the EC (or the representatives of lower-level legislative and regulatory action will therefore maintain an ongoing 'review and revise' strategy that will monitor and adapt to changes in AI use.

Regarding the current list of AI concerns, we draw attention to two academic papers that we have found useful in the understanding of societal AI risk; one written by the Centre for the Study of Financial Innovation (however the issues it raises are really industry-agnostic) [15], and another that is written by AI4People; a research initiative focussed on the foundations of a 'good AI society' [10]. Analysis of these two papers would suggest a wider set of risks than those currently listed in the EC's white paper.

Comparative Analysis (see Appendix A) between the EC' white paper and the CSFI's paper on *"the risks of AI in financial services"* [15] would infer five additional concerns:

- **Use of AI may lead to unmonitored decision making.**
- **Use of AI may incentivise users to act in ways that are not in the best interests of society.**
- **Use of AI without sufficient skills increase, will result in misuse of technology.**
- **Use of AI may result in less consumer choice.**
- **Use of AI will increase the complexity of technology systems, resulting in unexpected, unpredictable behaviour**

The "best interests of society", referenced in the CSFI paper, are the subject defined by Floridi et al in the AI4People paper [10].

- **Use of AI may devalue human skills.**
- **Use of AI may remove human responsibility.**
- **Use of AI may reduce human control.**
- **Use of AI may erode human self-determination.**

Question 6: Sufficiency of current legislation

Question 6. Do you think that the concerns expressed above can be addressed by applicable EU legislation? If not, do you think that there should be specific new rules for AI systems?

- ❑ Current legislation is fully sufficient
- ❑ Current legislation may have some gaps
- ✓ **There is a need for a new legislation**
- ❑ Other
- ❑ No opinion

Our response: Current legislation may have some gaps

The concerns about AI arise in a complex economic and social environment. While we anticipate that new legislation is required, it is important that the *need* (if properly defined before legislation) is considered. The EC should be setting out a strategic vision that addresses, in particular, its vision for 'a European Society', and what that means with respect to the concerns of AI, particularly those we emphasise above.

Question 6a. If you think that new rules are necessary for AI systems, do you agree that the introduction of new compulsory requirements should be limited to high-risk applications (where the possible harm caused by the AI system is particularly high)?

- ✓ **Yes**
- ❑ No
- ❑ Other
- ❑ No opinion

Our response: Regulation should focus on high risk applications

In principle, we support the limitation of regulation to 'high risk' applications, as we feel that regulation should be proportional to the consequences it seeks to control. Over regulation is likely to stifle innovation, however where there is a marketable claim to be made over 'AI safety' then this is a commercial decision, hence we also support a voluntary labelling system. However, the definition of high risk must be carefully defined (much more so than it is currently defined in EC literature). This must stem from what is considered 'safety'. The general product safety directive, 2001/95/EC [16], currently defines a 'safe product'[d] as "[…] consistent with a high level of protection for the **safety and health** of persons". It is this 'do no harm' principle that requires an AI-specific definition.

[d] 'product' is defined as "any product — including in the context of providing a service — which is intended for consumers or likely, under reasonably foreseeable conditions, to be used by consumers even if not intended for them, and is supplied or made available, whether for consideration or not, in the course of a commercial activity, and whether new, used or reconditioned".

Question 7: High risk applications

Question 7. If you wish, please indicate the AI application or use that is most concerning ("high-risk") from your perspective:

Our response: Risks should be described based on their application, not the technology itself

We observe that high-risk applications are not, typically, to do with the AI technology, but the context in which they are used. For example, consider facial recognition technology; the same technology may be considered a negative (in the use of such technology in police surveillance) or as a positive (when used within car safety systems to monitor driver alertness). Similarly, the same machine learning algorithm could be used in notably different sectors, from medical healthcare to transportation to military defence. Each will have different perspectives of what is considered 'high risk'.

While it is hoped that it is obvious, we emphasise the need for a regulatory framework where, for each identified 'high risk' application (within the various sector-contexts), that any data, AI or human-based decision-making system that support the high-risk application also subject to the high-risk regulations (i.e. requirements are 100% cascaded downstream from a top-level, industry specific application). Much in the same way that, to meet top-level safety requirements in the rail sector (e.g. avoid injury to life), all software contributing to that safety (e.g. the brake control software on a train, the indicators within signals) is validated to an agreed safety integrity level, Any software application contributing to an AI risk is equally validated under such scrutiny.

Our response: risks are best described as enterprise capabilities

We propose that 'high risk' applications are best defined through the terminology of *capability*. Dremel and Uebernickel [17] suggest that emergent AI and data technologies offer 14 new transformational capabilities to insurance enterprises. While some of the capabilities described appear relatively inconsequential to trust, a small handful pose risk to the protection of human civil liberties if implemented incorrectly.

- the exploitation of data for risk assessment and underwriting (TC2),
- the exploitation of data for claims handling (TC3),
- complimenting insurance services with prevention and recovery services (TC5), and
- offering risk-adjusted pricing (TC10).

These four capabilities all increase the level of intrusiveness of insurance services – through an increase in the amount of information known by the insurer, and the amount of influence over the customer's daily lives.

Question 8: Mandatory requirements

Question 8. In your opinion how important are the following mandatory requirements of a possible future regulatory framework of AI (section 5.d)?

	Not important at all	Not important	Neutral	Important	Very important	No opinion
The quality of training data sets					X	
The keeping of records and data					X	
Information on the purpose and the nature of AI systems					X	
Robustness and accuracy of AI systems					X	
Human Oversight					X	

Clear Liability and safety Rules					X	

Our response: All these aspects will need to be addressed.

Similar to our response to question 5, we perceive all of the above to be very important, as they represent various components of the same problem.

Question 9: Biometric Identification Systems

Question 9. In addition to the existing EU legislation, in particular the data protection framework, including the General Data Protection Regulation and the Law Enforcement Directive, or, where relevant, the new possibly mandatory requirements foreseen above (see question above), do you think that the use of remote biometric identification systems (e.g. face recognition) and other technologies which may be used in public spaces need to be subject to further EU-level guidelines or regulation:

- ❑ No further guidelines or regulations are needed
- ❑ Biometric identification systems should be allowed in publicly accessible spaces only in certain cases or if certain conditions are fulfilled (please specify)
- ❑ Other special requirements in addition to those mentioned in the question above should be imposed (please specify)
- ❑ Use of Biometric identification systems in publicly accessible spaces, by way of exception to the current general prohibition, should not take place until a specific guideline or legislation at EU level is in place.
- ❑ Biometric identification systems should never be allowed in publicly accessible spaces
- ✓ **No opinion**

Question 9a. Please specify your answer:

No opinion

Question 10: Voluntary Labelling Systems

Question 10. Do you believe that a voluntary labelling system (Section 5.G of the White Paper) would be useful for AI systems that are not considered high-risk in addition to existing legislation?
❑ Very much ❑ Much ❑ Rather not ❑ Not at all ✓ **No opinion**
Question 10a. Do you have any further suggestion on a voluntary labelling system?
No opinion

Question 11: Ensuring secure and trustworthy AI in respect of European values and rules

Question 11. What is the best way to ensure that AI is trustworthy, secure and in respect of European values and rules?
❑ Compliance of high-risk applications with the identified requirements should be self-assessed ex-ante (prior to putting the system on the market) ❑ Compliance of high-risk applications should be assessed ex-ante by means of an external conformity assessment procedure ❑ Ex-post market surveillance after the AI-enabled high-risk product or service has been put on the market and, where needed, enforcement by relevant competent authorities ✓ **A combination of ex-ante compliance and ex-post enforcement mechanisms** ❑ Other enforcement system ❑ No opinion ❑ Not at all
Question 11a. Please specify any other enforcement system:
See below
Question 11b. Do you have any further suggestion on the assessment of compliance?
See below

Our response: Both ex-ante and ex-post compliance methods are required

Firstly, it must be highlighted that, given the two types of ex-ante compliance presented, the inclusion of a combination option will not elicit true and accurate opinion on this issue. This aside, we do however contest that, due to the many dimensions of risk associated with the use of AI, that regulation cannot be achieved without a combination of ex-ante and ex-post measures. Distinction should also be drawn to the level of abstraction to which these measures are assessed. Within the context of the five response options, ex-post is described as a market-level assessment, looking for (e.g.) biases across the sector. While market-level surveillance is necessary to address societal consequences of AI, ex-post surveillance of individual organisations, and their AI-enabled products and services is also necessary.

Secondly, we would like to draw the readers attention to Commission Decision 2010/713/EU on modules for the procedures for assessment of conformity, suitability for use and EC verification to be used in the technical specifications for interoperability adopted under Directive 2008/57/EC of the European Parliament and of the Council. In this communication, a set of conformity 'modules' are described for the assessment of all aspects of a railway system related to the interoperability of the railway (read: safety). Much in the same way that

safety is assessed within the rail sector, the same principles of assessment can be applied to the AI disciplines. These conformity assessment modules separate design and production components of compliance. In general terms, manufacturers (read: companies that use AI) must either:

- assess each individual product (read: each instance of AI use), or
- prove the design of each product type (read: each type of AI use within an agreed context) and have in place a proven quality management system for production (read: each applied instance of AI use), or
- have proven quality management systems for both design and production (constrained based on application).

For all safety-related systems, compliance is assessed by an independent third party, who is accredited to conduct assessments following the criteria of the various conformity modules. As is common with all types of EU conformity assessment, the initial assessment (regardless of module) must be completed ex-ante. Where quality management systems are in place, these are assessed through regular surveillance audits (ex-post).

Further observations on Section 2

Our response: Trust is about more than regulation

As we introduced earlier, the trust is more than just regulation. Indeed, many researchers would question whether deterrent-based trust (i.e. regulation and sanction) fosters trust at all, or whether it is a substitute for trust [18]. Within section 5 of the white paper ("an ecosystem of trust"), it is implied that the only way that trust between consumers and those industries that use AI can only be achieved through regulation. While there is no commonly agreed definition of trust that spans every research discipline [18]–[22], all would concur that conformance to a set of rules is only one part of the equation.

This statement is not made to lessen the need for regulation of AI, IoT and associated data technologies, but to highlight the additional avenues available to the EC that can support increased trustworthiness, in parallel with a robust regulatory regime. A range of alternative modes can be considered, including information and education and incentive/market-based structures [23].

Of note, would be to develop a targeted education campaign, wider than Action 3 (Skills) of section 4 (ecosystem of excellence), which improves the level of education across the populous, rather than just those who enrol on AI-related educational courses. AI is a hyped and sensationalised subject matter, particularly within the general populous. We argue that societal acceptance of AI will be key to its long-term adoption success, and a key component of the trust relationship. For example, in one report [24] only 35% of insurance customers reported that they would be comfortable with businesses using AI to interact with them, and only 15% of customers would be comfortable with an insurance company using AI to monitor and analyse their daily activities. However, we can associate some of these opinions with 'fear of the unknown'. Indeed, the white paper acknowledges this in chapter 5: "lack of trust is a main factor holding back a broader uptake of AI". Most consumers do not understand what AI is and how it works. Consistently surveys are finding that consumers cannot identify products that they use daily that utilise AI technologies, or describe the characteristics of what AI should do [24], [25]. However, when the benefits and operation of these technologies are explained, consumers are much more accepting of its use [24], [25].

These activities should seek to tackle issues of AI hype, and awareness, while also improving the digital literacy of the populate to better protect society against emerging risks of digitisation, such as digital security and fake news.

Comments Section 3 – Safety and liability implications of AI, IoT and robotics

THE COMPANION PAPER – 'Report on the safety and liability implications of AI, the IoT and Robotics'

In the EC's "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics" [2] (a summary of which is provided on page 15 of the EC white paper on excellence and trust [1]), an analysis of the EU's legal framework w.r.t. AI use is conducted. It concludes that existing product safety legislation already supports an extended concept of safety, however it also argues that the explicit inclusion of new risks, caused by emergent digital technologies, could also provide additional clarity. Six key observations are noted:

- The autonomous behaviour of certain AI systems during its life cycle may entail important product changes having an impact on safety, which may require a new risk assessment. In addition, human oversight from the product design and throughout the lifecycle of the AI products and systems may be needed as a safeguard.
- Explicit obligations for producers could be considered also in respect of mental safety risks of users when appropriate (ex. collaboration with humanoid robots).
- Union product safety legislation could provide for specific requirements addressing the risks to safety of faulty data at the design stage as well as mechanisms to ensure that quality of data is maintained throughout the use of the AI products and systems.
- The opacity of systems based on algorithms could be addressed through transparency requirements.
- Existing rules may need to be adapted and clarified in the case of a stand-alone software placed as it is on the market or downloaded into a product after its placing on the market, when having an impact on safety.
- Given the increasing complexity of supply chains as regards new technologies, provisions specifically requesting cooperation between the economic operators in the supply chain and the users could provide legal certainty.

Additionally, the report identified that key characteristics of emerging digital technologies challenge aspects of existing liability frameworks, as the multi-actor nature of AI, IoT and robotic services makes it hard to trace damages back to a specific person or organisation. Two key requirements are highlighted:

- Persons having suffered harm caused with the involvement of AI systems need to enjoy the same level of protection as persons having suffered harm caused by other technologies, whilst technological innovation should be allowed to continue to develop.
- All options to ensure this objective should be carefully assessed, including possible amendments to the Product Liability Directive and possible further targeted harmonisation of national liability rules. For example, the Commission is seeking views whether and to what extent it may be needed to mitigate the consequences of complexity by adapting the burden of proof required by national liability rules for damage caused by the operation of AI applications.

Question 12: Emerging Product Safety Risks

Question 12. The current product safety legislation already supports an extended concept of safety protecting against all kind of risks arising from the product according to its use. However, which particular risks stemming from the use of artificial intelligence do you think should be further spelled out to provide more legal certainty?
✓ **Cyber risks** ✓ **Personal security risks** ✓ **Risks related to the loss of connectivity** ✓ **Mental health risks**
Question 12a. In your opinion, are there any further risks to be expanded on to provide more legal certainty?
See below

Our response: Clarification is always beneficial, provided it remains harmonious with existing legislation

Additional clarity is always beneficial. In response to question 5, we highlighted a select range of risks. Please see our comments against question 5 for further opinion on risk.

Question 13: Risk assessment procedures for products that change during their lifecycle

Question 13. Do you think that the safety legislative framework should consider new risk assessment procedures for products subject to important changes during their lifetime?
❑ Yes ❑ No ✓ **No Opinion**
Question 13a. Do you have any further considerations regarding risk assessment procedures?
No opinion

Question 14: Liability and compensation for AI applications

Question 14. Do you think that the current EU legislative framework for liability (Product Liability Directive) should be amended to better cover the risks engendered by certain AI applications?
✓ **Yes** ❑ No ❑ No Opinion
Question 14a. Do you have any further considerations regarding the question above?
See below

Our response: The study of systems of systems may help in the characterisation of a shared liability scheme

AI-enabled systems are, most often, systems of systems. Systems of systems (SoS) are characterised by the operational and managerial independence, and often geographical distribution of the constituent systems and system elements (that make up the SoS). They are also characterised by emergent behaviours caused by

the combined SoS and the evolutionary development processes that brought the SoS into being [26],[27]; SoS are normally complex constructs that emerge over time rather than being deliberately designed in a single process. This is the issue recognised in the issue of liability allocation. Within the insurance industry, this is further complicated, as this system of systems phenomena exists on two levels:

- As with all industries, the AI-enabled services are built upon the capabilities of a range technology suppliers to deliver the desired capabilities.
- Additionally, within insurance, the very concept of how risk is traded across the sector is also a system of systems problem. '*underwriters estimate customer's potential claims (losses) and decide whether to sell policies to them and at what price. Actuaries collect claim (loss) statistics and use this to calculate premium rates for different classes of policyholders. Insurance prices are determined by underwriters based on market conditions and the premium rates established by actuaries. Loss adjusters (claims adjusters) decide the value of claims when these are made by customers'* [Owadally et al., 2018]. '*In most insurance transactions, there is an intermediary, usually an insurance agent or broker, between the buyer and the insurer. In [many insurance] markets, the intermediary plays the role of "market maker," helping buyers to identify their coverage and risk management needs and matching buyers with appropriate insurers'* [Cummins & Doherty, 2006]. '*Other direct and indirect channels such as the contact centre [sic], internet, banks, aggregators [i.e. price and product comparison companies] and third-party retailers are part of the mix*' [Boobier, 2016].

  Resultantly, the relationship between the end customer and the person or organisation underwriting the risk is likely to be an indirect one, further increasing the change for error and the complexity of liability allocation.

A System of Systems (SoS) approach can help shape the language by which AI-enabled systems are described, and an analysis of appropriate liability. The four SoS archetypes (directed, acknowledged, collaborative and virtual) may help in the separation of different liability types. It may well be that identification of liability, at the resolution of individual or organisation, may not be possible or practical and instead, appropriate group-mechanisms are required for the recompense and rectification of fault. This is the idea of common enterprise liability [28], and the principle behind universal no-fault social insurance for AI related injuries [29].

Conclusion

In conclusion, we are happy to discuss any of the issues we raise further, please get in touch with techngi@lboro.ac.uk

Contributors:

Paul Timms MIET MINCOSE

Dr Melanie King

References

[1] European Commission, “White Paper on Artificial Intelligence - A European Approach,” 2020.

[2] European Commisson, “Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics.”

[3] European Commission, “The European Digital Strategy.” [Online]. Available: https://ec.europa.eu/digital-single-market/en/content/european-digital-strategy. [Accessed: 08-Jun-2020].

[4] European Commission, “A Europe fit for the digital age.” [Online]. Available: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en. [Accessed: 08-Jun-2020].

[5] European Commission, “Building Trust in Human-Centric Artificial Intelligence,” 2019.

[6] European Commission, “Stakeholder Consultation on Guidelines’ first draft .” [Online]. Available: https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/stakeholder-consultation-guidelines-first-draft. [Accessed: 08-Jun-2020].

[7] European Commission, “A European strategy for data,” 2020.

[8] European Commission and I. H.-L. E. G. on A. Intelligence, “A definition of AI: Main Capabilities and disciplines,” 2019.

[9] U. Von Der Leyen, “A Union that strives for more My agenda for Europe.”

[10] L. Floridi *et al.*, “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,” *Minds Mach.*, vol. 28, no. 4, pp. 689–707, 2018.

[11] A. Adler and M. E. P. Seligman, “Using wellbeing for public policy : Theory , measurement , and recommendations,” vol. 6, pp. 1–35, 2016.

[12] A. Milne and A. Zarifis, “TECHNGI Nov 2019 Conference Summary.” [Online]. Available: https://www.techngi.uk/2020/04/24/techngi-nov-2019-conference-summary/. [Accessed: 12-Jun-2020].

[13] European Commission, “Ethical Guidelines for Trustworthy AI,” *Futurium*, 2019.

[14] European Commisson, “Policy and Investment Recommendations for Trustworthy AI.”

[15] K. Patel and M. Lincoln, *It’s not magic: Weighing the risks of AI in financial services*. Centre for the Study of Financial Innovation (CSFI), 2019.

[16] “2001/95/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 3 December 2001 on general product safety (Text with EEA relevance).”

[17] E. Stoeckli, C. Dremel, and F. Uebernickel, “Exploring characteristics and transformational capabilities of InsurTech innovations to understand insurance value creation in a digital world,” *Electron. Mark.*, vol. 28, no. 3, pp. 287–305, 2018.

[18] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, “Not so different after all: A cross-discipline view of trust,” *Academy of Management Review*, vol. 23, no. 3. pp. 393–404, 1998.

[19] J. K. Butler, “Toward Understanding and Measuring Conditions of Trust: Evolution of a Conditions of Trust Inventory,” *J. Manage.*, vol. 17, no. 3, pp. 643–663, Sep. 1991.

[20] M. Tschannen-Moran and W. K. Hoy, “A Multidisciplinary Analysis of the Nature, Meaning, and Measurement of Trust,” *Rev. Educ. Res.*, vol. 70, no. 4, pp. 547–593, Dec. 2000.

[21] G. Dietz and D. N. Den Hartog, “Measuring trust inside organisations,” *Pers. Rev.*, vol. 35, no. 5, pp. 557–588, 2006.

[22] B. G. Robbins, “What is Trust? A Multidisciplinary Review, Critique, and Synthesis,” *Sociol. Compass*, vol. 10, no. 10, pp. 972–986, Oct. 2016.

[23] "Using alternatives to regulation to achieve policy objectives," 2014.

[24] Pega, "What Consumers Really Think About AI : A Global Study Executive summary : The AI / consumer paradox," 2017.

[25] ARM and North Star, "AI Today, AI Tomorrow."

[26] B. of K. and C. to A. S. E. (BKCASE), "SEBoK," *SEBoK v2.0*, 2019. [Online]. Available: https://www.sebokwiki.org/wiki/Guide_to_the_Systems_Engineering_Body_of_Knowledge_(SEBoK). [Accessed: 29-Jul-2019].

[27] I. 21841:2019(E), "Taxonomy of systems of systems," in *Systems and Software Engineering*, First Edit., 2019.

[28] H. R. Sullivan and S. J. Schweikart, "Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI?," *AMA J. Ethics*, vol. 21, no. 2, pp. 160–166, 2019.

[29] J. Yoshikawa, "Sharing the Costs of Artificial Intelligence: Universal No-Fault Social Insurance for Personal Injuries.," *Vanderbilt J. Entertain. Technol. Law*, vol. 21, no. 4, pp. 1155–1187, 2019.

[30] K. Martin, K. Shilton, and J. Smith, "Business and the Ethical Implications of Technology: Introduction to the Symposium," *J. Bus. Ethics*, vol. 160, no. 2, pp. 307–317, 2019.

Appendix A – Analysis of EC's AI Concerns versus those raised by the CSFI

EC white paper of "AI concerns"	CSFI paper [15] "risk drivers" and "key risks"
AI may endanger safety	Pressure to move too fast: Pressure to deploy AI solutions quickly to remain competitive may lead to risks, including insufficient testing and an overreliance on AI specialists.
	Changing incentive structures: The benefits to successful actors and the risks of getting left behind create powerful incentives for firms to collect data and implement AI solutions on a rapidly accelerated timeline.
	New regulatory challenges: AI poses new challenges for regulators and policymakers because of its technical complexity, the ethical questions it raises and its potential to fundamentally transform market structures.
AI may breach fundamental rights (such as human dignity, privacy, data protection, freedom of expression, worker's rights, etc.)	Perverse behaviour of AI models: AI models can lead to the propagation of biases that can be very difficult to identify and root out. They can also perform poorly in previously unencountered situations.
AI is not always accurate	
The use of AI may lead to discriminatory outcomes	Optimisation at the expense of social benefits: AI enables institutions to evaluate risks at a much more granular level, which could disadvantage certain customers and challenge conceptions of fairness.
	Over-reliance on AI: Resources could be wasted on AI if it is implemented 'for its own sake', or if the people reliant upon it are unable to interpret or work with their outputs effectively.
AI may take actions for which the rationale cannot be explained	Opacity and complexity: A trade-off at the heart of many AI models is that, generally speaking, the more effective the algorithms, the more difficult they are to scrutinise. Insufficient transparency: The difficulty of understanding and explaining decisions made or augmented by AI could damage trust in financial services.
Use of AI may lead to unmonitored decision making	Distancing of humans from decision making: AI is different from previous forms of automation because it enables many actions to be taken without explicit instructions from humans.
Use of AI may incentivise users to act in ways that are not in the best interests of society	Data acquisition and aggregation: Using AI creates strong incentives for financial institutions to collect, aggregate and centralise data, increasing concerns about data security and privacy.
Increased use of AI, without sufficient skills increase, will result in misuse of technology.	Talent gap: There is an acute shortage of specialists who can design, develop, deploy, test and maintain AI systems – particularly of those who have knowledge of financial services.
	Knowledge gap and unrealistic expectations: AI systems could fail spectacularly if decision-makers who don't understand the

	technologies do not set appropriate expectations or provide AI teams with the right resources.
Use of AI may result in less consumer choice	Market concentration: While AI has spurred competition, it may also lead to more market concentration and the erection of barriers to entry since its 'winners' benefit from economies of scale and powerful new network effects.
Use of AI will increase the complexity of technology systems, resulting in unexpected, unpredictable behaviour	Increased interconnectedness: Use of AI might create new kinds of interconnectedness in financial markets – at the data level, the IT systems level, and the decision-making level.