Reaction Document to the Public Consultation on Malta's Ethical AI Framework

6th September 2019

MALTA IT LAW
ASSOCIATION

Malta Information Technology Law Association (MITLA)

Authors:
Dr Antonio Ghio
Dr Gege Gatt
Dr Lena Sammut
Dr Deo Falzon
Dr Yanica Sant

www.mitla.org.mt
info@mitla.org.mt

Table of Contents

Introduction

As a voluntary organisation founded with the aim of promoting and actively participating in the advancement of information technology law in tandem with developments in technology, MITLA brings together a distinctive mix of voices ranging from the relevant industries, legal practice, regulatory bodies and academia. For this reason, its input in the development of a comprehensive artificial intelligence (hereinafter "AI") strategy for Malta is highly beneficial. In order to make the development and optimisation of AI a success story for Malta, the ethical building blocks upon which such an AI framework is structured must echo the ethical principles espoused at European level and internationally, while also taking into account the criticism which has been levied at same in order to avoid their shortcomings.

MITLA's response takes into account the principal reactions to the High Level Expert Group (hereinafter "HLEG") for Trustworthy AI set up by the European Commission and the Guidelines published by same, since as disclosed in the consultation document published by the Malta AI Taskforce (hereinafter the "Consultation Document"), this was the biggest singular influence leading to the drawing up of the Consultation Document. It continues by considering the societal impact of AI and the ethical principles which should be espoused to ensure that AI is not self-serving, but serves and furthers the interests of humanity, together with the legal framework which is essential for the adoption of an AI strategy which is trustworthy by design. The third chapter focuses on the relationship between industry and ethical AI. The reaction document finally takes on a comparative exercise of the ethical efforts which have been undertaken in the United Kingdom in the field of AI.

MITLA's aim in providing its reaction to the Taskforce's Consultation Document is to further the discussion on Malta's ethical AI framework in a constructive, reasoned and impartial manner.

Executive summary

Embedded in this reaction document are a number of proposals and suggestions which MITLA wishes to put forward for consideration in creating an ethical AI framework for Malta.

Key among these observations are:

(a) A recommendation in chapter 1 to place greater emphasis on the creation and bolstering of legal mechanisms which seek to address and redress the intrinsic biases of AI, being that they are the subject of human input, in order to ensure a human-centric approach, rather than on the use of the term 'trustworthiness', both because of the different social and ethical connotations of the term, as well as because as MITLA opines in line with various commentators referred to below, the trustworthiness of AI is not to be derived from the systems which implement it.

If, however, the use of trustworthiness as a metric against which to measure the suitability of AI to contemporary and future society is not appropriate, what then should form the basis of a truly ethical AI framework? The principles which MITLA proposes should replace or at least supplement the moniker of trustworthiness are outlined in chapter 2 of this document.

(b) MITLA's principal assertion is that prior to engaging in any major developments and/or amendments of the existing legal framework, digital rights should first be introduced into the Maltese Constitution. This would put Malta at the forefront in the movement towards building an ethical AI framework and would secure rights which are becoming ever more important in contemporary society and will continue to gain importance for future generations, including the right to informational self-determination, the right to unhindered development of personality, the right to access to truthful sources of information balanced against freedom of expression and the potential role of Government in deterring the spreading of false or misleading information and the provision of guidelines for the use of AI systems by which potentially harmful content may be flagged and dealt with.

(c) MITLA calls for a systemic analysis of the legislative needs required for the creation of new liability mechanisms or the adaptation of existing legal

liability models to cater for cases of harm caused by AI stakeholders operating an AI system to another party.

(d) Adding to the Consultation Document's recommendation regarding accountability, MITLA further recommends that in order to counter for a potential malfunctioning or worse of AI systems, AI systems should have built-in 'kill-switches' making it easy for humans to manually override, take over and redact any potential unwanted consequences of these AI systems.

(e) MITLA makes a strong recommendation to Government not to grant distinct legal personality to AI systems, as doing so would undermine the fundamental principle of human-centric AI, AI which respects fundamental human rights and accountability in AI.

(f) Any local legislative efforts towards the regulation of AI should take into account previous and current European and global efforts for coordination, as otherwise, Malta runs the risk of falling out of step with legal mechanisms tackling the use of AI. At the same time, Government has an equal obligation to ensure that AI systems adhere to and reflect the ethical and moral principles of Maltese society.

(g) It is MITLA's recommendation that local competition law enforcement authorities should proactively monitor progress in AI technology and the use and commercialisation of big data. Tied to this, Government should undertake regular reviews of competition law to ensure that consumers stand to gain from the development and deployment of AI systems. In tandem, Government should create the infrastructures necessary to promote open access to datasets for academic research and/or non-commercial use via open source or other similar licensing arrangements. Data held by public sector bodies, private sector bodies and industry organisations should be accessible and open, as appropriate and where possible.

(h) MITLA proposes that the development, deployment and use of AI systems should require the appointment of an AI Ethics Officer, whose role would mimic the responsibilities and functions of the Data Protection Officer under the General Data Protection Regulation.

(i) While recognising that current local, European and international IP do not adequately cater for AI-enabled and/or AI-created IP works, MITLA recommends that Government considers and actively monitors the

development of legislation in this regard which provides adequate protection while, as much as possible, not hindering the flow of data within and the development of AI systems.

(j) On the subject of data agency, it is time for the industry and also legislators to recognise that digital consent is not serving the purpose of informing citizens about the use of their data, including the extent thereof and their rights attached thereto. A new and effective way of informing and engaging data subjects must be found.

(k) AI must not be used for the sake of itself, but for the sake of society - industry must provide evidence of the advantages to be gained by adopting an AI system and its fitness for the purpose at hand.

(l) Another obligation which MITLA proposes to be expected of the industry is that of always being in a position to trace back the origins and logic behind a particular decision or outcome, so as to counter the dangers of machine learning and bias.

(m) From an academic perspective, modules on ethics should be form an intrinsic part of, and integrated within any science, technology, engineering and mathematics programme of studies and not be merely considered as an after-thought.

(n) Government should allocate resources and create incentives to bring together different stakeholders in the field of AI, including engineers and designers, scholars and social scientists.

(o) Finally, Government should invest its finances and resources in securing and protecting its data by actively participating in efforts on a European level to address Europe's current dependence on data centres located outside the European Economic Area.

Chapter 1: Examining the sources of the Consultation Document

This section attempts to provide some visibility on how commentators have responded to concepts found within and those leading up to the HLEG on Artificial Intelligence Guidelines, which admittedly is the biggest singular influence on the Ethical Framework consultation document published by the Malta AI taskforce in August 2019. Since the Maltese document is intrinsically based on such guidelines it naturally follows that criticisms made to the EU guidelines would also apply to the Maltese approach and help fine-tune it.

The importance of conducting a widespread and serious ethical discussion about AI cannot be overstated; indeed, some commentators and futurists are referring to as a development affect all aspects of society which will introduce Life 3.0.[1] The "Trustworthiness" moniker in AI has been set high enough on most AI agendas and seems cemented as a term towards which entities developing AI frameworks should aspire to. The High Level Expert Group on Artificial Intelligence (hereinafter "HELG") has gone as far as to include it as an intrinsic part of ethics in AI and other jurisdictions are following the example and including "trustworthiness" as the crowning feature of an AI which is seemingly ethical in nature.

Some commentators are putting forward the argument that the way trustworthiness is being shoehorned in the AI discussions, is essentially flawed[2].

---

[1] Max Tegmark, Life 3.0: Being Human in the Age of Artificial Intelligence (Knopf 2017). (Life refers to the difference stages of human life starting from inception, Life 1.0: Biological Origins, Life 2.0: Cultural Developments and Life 3.0: Technological Age of humans.)

[2] *"While we agree that trustworthiness is a key objective for any system, the HLEG must also acknowledge the limitations of current methods for mitigating bias in machine learning models. In many contexts and applications, truly trustworthy AI remains hypothetical."*

Center for Democracy and Technology, 'Comments of The Center for Democracy & Technology on European Commission's High Level Expert Group on Artificial Intelligence (AI HLEG)'s Draft Ethics Guidelines for Trustworthy AI' (2019) <https://cdt.org/files/2019/02/comment_-EU-Commission-HLEG-AI-guidelines-1.pdf> accessed 28 August 2019

Combining trustworthiness with "doing good" may be muddying the waters, since an AI system can only subscribe to doing "good" if it is programmed to achieve a "good" result[3]. The definition of good may vary somewhat according to the applicable anthropological and sociological interpretations. "Trust" encompasses a collection of verbs and adjectives such as "confidence" and "reliance" which seek to define the outsourcing of care and its placing within an external entity or system whether scientifically measurable, or otherwise. Trustworthiness does not necessarily entail a positive result for the majority or the minority, and therefore, many positions advocate that AI systems defined as trustworthy should be separated from a "do good" expectation or discussion and rather, should focus on predictability, explainability and a clear line of custody and chain of events which lead to transparency and bias auditability when affected by human intervention. The principles which MITLA proposes should replace or at least supplement the moniker of trustworthiness are outlined in Chapter 2 of this document.

The varying results of preferable choices which an AI system can take is interestingly portrayed through the open test developed by Massachusetts Institute of Technology (hereinafter "MIT"), the Moral Machine[4]. The test presented two polarising scenarios in which the user is meant to judge between two decisions taken by the AI system loaded in a self-driving vehicle in equivalent scenarios with two diametrically opposed results. The ethical ramifications of the results, which vary greatly from one country to another as well as between social groups[5] portray the difficulty and danger in linking trustworthiness to an obligation to "do good".

Chronology of development of the discussion at EU level and commentary thereon

On the 9th March 2018, the European Group on Ethics in Science and New Technologies (European Commission) (hereinafter "EGE") issued a statement on

---

[3] Ibid.

[4] Scalable Cooperation, 'Moral Machine' <http://moralmachine.mit.edu/> accessed 28 August 2019.

[5] Ibid.

Artificial Intelligence, Robotics and Autonomous Systems,[6] flagging the fact that new AI systems based on a version of machine learning referred to as "deep learning" are increasingly opaque and it is increasingly difficult to ascertain and decipher their results beyond the initial algorithms they would have been programmed with. As deep learning builds on its own biases and weights which may have been present in its past iterations, these become ingrained in the system and this problem will exacerbate as the program progresses, unless corrected. In essence, AI through machine learning (especially deep learning) fuelled by raw big data is becoming more powerful as it grows harder to track its inner workings. These observations raise questions about safety, prevention of harm and risk mitigation to make AI devices secure, the need to introduce moral responsibility (liability), governance and testing and to make AI safe for society as well as to ensure democratic decision-making, explicability and transparency.

On the 25th April 2018, the European Commission issued a Communication on Artificial Intelligence for Europe[7] which effectively set out the need to ensure an appropriate ethical and legal framework, based on the Union's values and in line with the Charter of Fundamental Rights of the EU and sets the way forward for the development of AI ethics guidelines having regard to the aforementioned EGE. The Maltese position seems well aware of the Commission's publications but seems to have lost sight of having regard to the ethical base as advocated by the EGE's positions. MITLA advises that any ethical discussions should consider the positions advocated by the EGE especially issues of justice, equity, solidarity (the fair distribution of benefits and equal opportunities) as well as sustainability, amongst others.

On 19th December 2018, the EGE published a document concerning the Future of Work, Future of Society[8] which presents certain red-line areas of AI development and reinforces non-bias issues. It also comments that the industry

---

[6] European Group on Ethics in Science and New Technologies, 'Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems' (Brussels, 9 March 2018) <http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf>.

[7] European Commission, 'Artificial Intelligence for Europe' (Communication) COM (2018) 237 final.

[8] European Group on Ethics in Science and New Technologies, 'Future of Work, Future of Society' (2018).

in general proposes a solution to people replaced by AI by forcing them to upskill or re-skill, but there is little mention of what would happen to the low skilled or those incapable of upskilling.

MITLA notes that on a similar vein, the discussions in Malta concerning AI have so far failed to address these concerns and these run the risk of not being addressed at all if they are not brought to light at this initial stage.

On the 29th January 2019 the EGE published an Open Letter to the Commission drawing attention to shortcomings in the process of the HELG.[9] The Open Letter flags serious concerns regarding the approach, societal vision and ethical reasoning behind the AI ethics being developed:

1. Technological mastery seems to be an end in and of itself, where ethical and social values are termed as a given as long as they do not hinder technological progress;

   MITLA posits that human-centric AI should not just focus on wellbeing but must also be based on human dignity. This level of detailed discourse is also missing from the Maltese approach till now.

2. The EU's overarching framework of human rights should not be abstracted in a "do-good" mantra without delving into what the good is and how to identify it. Ultimately, MITLA would also comment that a "do-good" mantra, despite the best of intentions, would not replace the necessity of a clear chain of command in events performed through AI intervention. Despite the various complexities it may feature, (discussed above), AI ultimately remains a set of commands which are executed and without proper understanding of meaning. The Turing and Chinese Room tests clearly showcase the fact that despite the progress and despite the enormous number of permutations which may be coded in the AI as well as those which can be obtained through Machine Learning, deep learning would still depend on a series of biases introduced by human handlers.

---

[9] Open Letter by the European Group on Ethics in Science and New Technologies to the European Commission (29 January 2019) <https://ec.europa.eu/info/sites/info/files/research_and_innovation/ege/ege_ai_letter_2019.pdf> .

3. The current approach of conflating legal obligation without considering compliance and enforcement is seen as ethics washing.

During February 2019, the Centre for Democracy & Technology (hereinafter "CDT") in Washington commented on the HLEG draft guidelines which have since been supplemented with a follow-up publication following the final HLEG guidelines.[10,11] The CDT agrees that

"trustworthiness is a key objective for any system but the limitations of current methods in mitigating bias in machine learning models makes truly trustworthy AI merely hypothetical and therefore it can be argued that terming AI as trustworthy would merely constitute an ethics-washing approach. Moreover, framing ethical AI as trustworthy AI would also be humanizing it and attributing a quality which should not be an end goal for AI".

MITLA recommends that this personification of AI should be avoided in legal and ethical discussions as it risks attributing values to a sensitive discussion which is already polarising human sentiment due to its societal implications.

Moreover, MITLA posits that trustworthy AI is not achieved on the ethics or technical robustness of the AI implementation *per se*, but on the robustness and appropriateness of the legal system within which an AI application sits. Greater emphasis is to be placed on mechanisms and processes meant to challenge AI and assessing the entire systems including policies, laws and human technology interactions.

The CDT continues that –

---

[10] Stan Adams and Natasha Duarte, 'Just Be Ethical: High Level Guidelines on AI are Fine, but Offer Little Guidance - Center For Democracy & Technology' (2019) <https://cdt.org/blog/just-be-ethical-high-level-guidelines-on-ai-are-fine-but-offer-little-guidance/> accessed 28 August 2019.

[11] For an interesting outline of major incidents reflecting the pitfalls in Machine Learning Bias consult page 4-6 of Anastasia Siapka, 'The Ethical And Legal Challenges Of Artificial Intelligence: The EU Response To Biased And Discriminatory AI' (Postgraduate, Panteion University of Athens, Panteion University of Political and Social Sciences 2018).

"When fundamental human rights are translated to ethical "principles and values" to govern AI, it is likely that different stakeholders and decision makers will apply the principles differently. This is particularly true of beneficence ("do good"). We have found that between civil society and industry, and even among industry actors, beliefs about what technologies or designs benefit society can diverge widely."[12]

MITLA believes that effective methods of redress are to be given due prominence, as will be expounded in chapter 2 of this reaction document.

On bias and its elimination, as identified in the HLEG document, such principle-based approach should be avoided since bias is a result of society and therefore as a reflection of society cannot be eliminated from AI if society itself, which is analysing that AI, incorporates the bias itself.[13]

MITLA notes that the Consultation Document refers to a series of absolutes when referring to the elimination of bias which should preferably refer instead to the inclusion of biases which would be incompatible with normal rules of law in any other human endeavour unrelated to AI and therefore covering discriminatory, prejudicial bias, etc.

On 8th April 2019 the HELG issued the final version of the Ethics Guidelines for Trustworthy AI.[14] Again, a common thread among critics and commentators is the concept that sociological robustness of AI ethical principles, rather than just the technological robustness, is to be upheld.[15] Considering that ethical principles depend greatly on contextual realities and cultural differences, consensus on ethics may well prove too difficult if realistically sought for, and it

---

[12] Ibid.

[13] Joanna Bryson, 'My Comments/Critiques on The EU's High Level Expert Group on AI's "Ethical Guidelines"' <https://joanna-bryson.blogspot.com/2019/02/my-commentscritiques-on-eus-high-level.html> accessed 28 August 2019.

[14] Independent High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (European Commission, 2019).

[15] Roel Dobbe, 'Clarifying Confusion Around the EU Draft Ethics Guidelines for AI' (*Medium*, 2019) <https://medium.com/@roeldobbe/clarifying-confusion-around-the-eu-draft-ethics-guidelines-for-ai-5e0505bf336b> accessed 28 August 2019.

is in fact argued by some commentators that AI should instead be democratically legitimate and the crossroads between intellectual property (hereinafter "IP") existing in an AI system and the need for accountability is to be better explored, which discussion is completely missing from the Consultation Document. MITLA's proposals on the manner in which IP existing in an AI system should be managed from a legal point of view are put forward in chapter 2 of this reaction document.

It is a given that as any other invention or product that is made available, AI should be compliant with fundamental human rights. Plenty of discourse has identified a human-centric guided approach to Trustworthy AI. In parallel however a human-centric approach should be aware that: "*Eventually, the epistemic and normative limitations of AI should be acknowledged, leading to its use for facilitating instead of replacing human decision-making.*"[16]

One should similarly heed the criticism being thrown against the HELG itself from within its ranks mainly in relation to the removal of the Red Line wording in relation to the use of AI that people cannot understand, social scoring and lethal autonomous weapon systems, which "Red Line" wording has now been replaced by the softer sounding title: "Critical Concerns". In an exceedingly honest publication, Thomas Metzinger, Professor of Theoretical Philosophy representing the European University Association, criticises some Ethics Committees, referring to them as setups resorted to by politicians when the complexity of the subject matter faced is simply too much to comprehend[17]. He also discourages following guidelines implemented by states such as China which are seemingly progressing full steam on AI when their totalitarian status poses a fundamental problem with genuine ethics. This poses greater pressure on the EU and its member states to bear the responsibility of true ethical guidance in view of AI. Ultimately, he advocates for society to take concrete hold of the process away from the industry stakeholders to ensure that the fundamental building blocks are kept in line with fundamental values.

---

[16] Siapka (n 10) [41].

[17] Thomas Metzinger, 'Ethics Washing Made In Europe' (*Der Tagesspiegel*, 2019) <https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html> accessed 28 August 2019.

MITLA notes that the Consultation Document avoids issues such as education and labour impact as these will be addressed in a wider AI strategy. MITLA's proposals on the importance of including ethics modules in Science, Technology, Engineering and Mathematics (hereinafter "STEM") programs from an early stage into and throughout academic curricula are presented in chapter 3 below.

In conclusion, MITLA embraces the advice given by various critics to frame ethical discussions within sociological realities, as an approach which fails to do so would not reflect a true human-centric and ethical approach but rather a personification of largely obtuse adjectives which are meaningless and useless in the face of the change being brought forth by AI developments and the sociological components that face the greatest risk. To paraphrase the anthropologist S.A. Applin, if we want ethical discussions concerning AI, the discussions should involve the groups and cultures they affect[18].

---

[18] Sally A. Applin PhD, 'Everyone's Talking about Ethics in AI. Here's what they're Missing' (*Fast Company*, 2019) <https://www.fastcompany.com/90356295/the-rush-toward-ethical-ai-is-leaving-many-of-us-behind> accessed 28 August 2019.

Chapter 2: Ethical purpose and social impact of AI

For the purposes of this section of the response document, MITLA has based its response using the Global Policy Framework for Responsible AI published by the International Technology Law Association (hereinafter "ITECHLAW"), and has endeavoured as much as possible to align the said Global Policy Framework with the proposed principles and guidelines contained in the Consultation Document. In light of the comment contained at Pg. 5 of the Consultation Document, any discussion on the impact of AI and the Transformation of Work will be addressed at a later stage.

MITLA fully subscribes with the eight (8) high level principles adopted by ITECHLAW within its Global Policy Framework for Responsible AI which are further expounded upon in the following sub-sections.[19] MITLA recognises that most of these principles have been reflected in some form or manner in the Consultation. However, particularly in the level of importance placed on human rights and their role in ethical AI, MITLA believes that any proposal for Malta's Ethical AI Framework should encapsulate these eight principles more comprehensively.

ITECHLAW's Global Policy Framework for Responsible AI – the eight principles

Principle 1: AI - Ethical Purpose and Societal Benefit

Any development, deployment or use of AI systems should do so in a manner compatible with human agency and the respect for fundamental human rights (including freedom from discrimination). At this early juncture, MITLA strongly believes that the proposed introduction of Digital Rights in the Maltese Constitution should take centre-stage and this not only vis-à-vis the discussion on AI through this Consultation Document but especially in the wider landscape of a constitutional review.

---

[19] The reader is invited to refer to the original document, inclusive of its sources, published by ITECHLAW available at <https://www.itechlaw.org/ResponsibleAI> accessed 5 September 2019.

On this point one would wish to refer to the White Paper published by the Government of Malta in 2012[20] but most importantly to the Private Members' Bill presented in May 2014 by which the proposal for the introduction of these digital rights was not only limited to the insertion of guiding declaratory principles (as unenforceable rights) but through their inclusion in Chapter IV of the Constitution.

MITLA proposes that key in this discussion is the introduction of the right to informational self-determination (known in German constitutional doctrine as *Informationelle Selbstbestimmung*) and which should also serve an important metric with which any processing of data (whether or not carried out by AI systems) should be measured. Closely associated with this, a discussion on the introduction to the right to the unhindered development of personality should be initiated.

Any development, deployment or use of AI systems should be accompanied by the monitoring of the implementation of such AI systems and any organisation which is developing, deploying or using AI systems should act to mitigate against consequences of such AI systems (whether intended or unintended) that are inconsistent with the ethical purposes of beneficence and non-maleficence. Any organisation involved in the AI landscape (being designer, developer and/or user) should assess the social, political and environmental implications of such development, deployment and use in the context of a structured Responsible AI Impact Assessment that assesses risk of harm and, as the case may be, proposes mitigation strategies in relation to such risks.

Furthermore, organisations that develop, deploy or use AI systems to filter or promote informational content on internet platforms that is shared or seen by their users should take reasonable measures, consistent with applicable law, to minimise the spread of false or misleading information where there is a material risk that such false or misleading information might lead to significant harm to individuals, groups or democratic institutions.

---

[20] White Paper, 'Introduction of "Digital Rights" in the Constitution of Malta' (2012) <https://mita.gov.mt/en/News/Documents/1_34533%20MITC%20White%20Paper%20Doc%20A4%20web%20final.pdf> accessed 5 September 2019.

This specific point has specific resonance in Malta and, MITLA believes, has not been given its due importance in the Consultation Document, whereby recent years saw the meteoritic rise in the spreading of false or misleading information on the internet.

Another level of this discussion should also factor what (if any) role should the State play in regulating such use of AI and this to ensure that our fundamental right to receive, share and impart information (including online) is properly safeguarded. AI has the potential to assist in efficiently and proactively identifying (and, where appropriate, suppressing) unlawful content such as hate speech or weaponised false or misleading information. Therefore, AI research into means of accomplishing these objectives in a manner consistent with freedom of expression should be encouraged.

Moreover, AI systems on platforms deployed to filter or promote informational content that is shared or seen by their users should provide a mechanism by which users can flag potentially harmful content in a timely manner as well as mechanisms by which content providers can challenge the removal of such content by such organisations from their network or platform in a timely manner.

In this sense, MITLA proposes that Government should provide clear guidelines relating to AI systems on platforms targeting and monitoring content and how to identify prohibited content in full respect, both of the rights to dignity and equality and the right to freedom of expression. In any event, it is the courts which should remain the ultimate arbiters of lawful content.

Principle 2: AI – Accountability

The development, deployment or use AI systems and any national laws that regulate such use should respect and adopt all applicable principles for responsible AI. In all instances, humans should remain accountable/responsible for the acts and omissions of AI systems as it should be the organisations (i.e. humans) that develop, deploy or use AI systems that ought to be accountable for the consequences of their actions. Such organisations should not be able to elude responsibility for legal prejudices or ethical faults, for property damage or for human injury that is caused by AI, regardless of the degree of AI autonomy by simply blaming the machine. Accountability should thus be understood and

applied in such a way as to always keep humans in a position to modify, monitor or control AI systems, and as such, be accountable for them.

Furthermore, if one expects AI stakeholders to respect and implement principles of transparency and explainability, fairness and non-discrimination, safety and reliability as well as of fair competition and face consequences when they fail to do so, relevant actors need to be given the information necessary for them to respect their obligations and navigate successfully within this new ecosystem. It needs to be underlined that it should be the legal responsibility of AI stakeholders to operate AI systems in a way that respects, at all times, existing laws—whether fundamental, penal or civil—and regulations.

Similarly, liability mechanisms will equally apply in the case where an AI stakeholder operates an AI system and causes injury to another party. One of the pillars of any legal system is its liability framework—as it ensures the trust of society, by ensuring compensation and justice for the damage caused when a person does not respect the law.

Various legislations enacted or proposed throughout the world attempt at solving these important challenges. These include the Canadian Personal Information Protection and Electronic Documents Act[21], the proposed Algorithmic Accountability Act[22] in the US or the proposed Model Artificial Intelligence Governance Framework[23] in Singapore.

So far, MITLA notes that Malta has not come up with legislative efforts of its own to tackle this important facet of AI framework development.

---

[21] Personal Information Protection and Electronic Documents Act (S.C. 2000, c. 5) <https://laws-lois.justice.gc.ca/ENG/ACTS/P-8.6/index.html> accessed 5 September 2019.

[22] Algorithmic Accountability Act of 2019 <https://epic.org/privacy/policy/Algorithmic-Accountability-Act-2019.pdf> accessed 5 September 2019.

[23] Personal Data Protection Commission of Singapore, 'A Proposed Model Artificial Intelligence Framework' (January 2019) <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf> accessed 5 September 2019.

In certain cases, the use of AI systems by public institutions will be subject to the same accountability standards and obligations as private actors. However, in others cases, especially those involving police, judicial, administrative or military institutions, the use of AI systems by governments should be subject to an even higher standard of conduct, to account for the importance of the rights involved and the risk that they be harmed. Notably, governments must be particularly attentive to their constitutional and international obligations, including respect for fundamental human rights. Governments should also be accountable to their constituents for the enactment of well-designed and effective legislation and regulations on AI. As such, governments are not only developers, deployers and users of AI systems, but are also the ultimate regulators of the activities of other stakeholders.

MITLA subscribes to the position that governments should not work on an overarching "law of AI" Existing regulators and governing bodies are in the best position to make such amendments to regulations or to create standards in their respective fields, as opposed to an overarching authoritative body for "all AI," which, although interesting in theory, will most certainly prove difficult in application, given the various fields that AI touches upon. It is possible however that in some cases, AI-specific regulation may be warranted for a given sector to ensure stakeholder accountability.

MITLA further recommends that in order to ensure that AI systems can be controlled in case they malfunction or 'go rogue,' especially AI robots, developers should build 'kill-switches' and enable easy manual override to protect against unwanted consequences.[24] Instant switch-off functionality is paramount because reinforcement-learning systems may find ways to cut out the operator from the loop.[25] Some scholars have even argued that developers/manufacturers must provide "programmatic-level accountability" by being able to prove why a system operates in certain ways, in order to address

---

[24] European Commission, *European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics,* (2015/2103(INL)) at 59 (f) http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN#BKMD-12 accessed 5 September 2019.

[25] BBC, "Google Developing Kill Switch for AI" (8 June 2016) <https://www.bbc.com/news/technology-36472140> accessed 5 September 2019.

legal issues and apportion culpability.[26] Such measures that ensure programmatic/procedural accountability can be mandated for developers who are engaged in developing high-risk AI systems.[27]

Good AI governance will also include making sure the stakeholder is complying with all applicable laws and regulations. Stakeholders must be confident that their AI systems respect applicable rules.

MITLA recommends the appointment of a Chief AI Officer and the creation of an AI committee to help with compliance efforts as one of their tasks should be to monitor the evolution of AI ethical and legal obligations and inform the organisation about new requirements. More suggestions on the manner in which industry players should ensure the ethical development and implementation of AI systems are made in Chapter 3 below.

Beneficial AI demands human accountability. General principles, even if well-intended, are useless without enforceable accountability regimes and without efficient governance models.

With respect to AI Accountability, the designation of individuals accountable for the organisation's compliance with the AI principles is paramount as well as establishing when an AI impact assessment is required, the structure and format of carrying out same, reporting lines, the implementation of 'Responsible AI by Design' principles within the product lifecycle, as well as ongoing training and awareness. The intensity of the accountability obligation will vary according to the degree of autonomy and criticality of the AI system. The greater the level of autonomy of the AI system and the greater the criticality of the outcomes that it may produce, the higher the degree of accountability that will apply to the organisation that develops, deploys or uses the AI system.

---

[26] Institute of Electrical and Electronics Engineers Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design, First Edition" (2019) <https://ethicsinaction.ieee.org> accessed 5 September 2019.

[27] Further guidance could be sought from projects like the Directive on the Use of Machine Learning for Decision-Making[27] recently published by the Government of Canada as well as the proposed model for Artificial Intelligence Impact Assessment (AIIA) issued by the Dutch Platform for the Information Society <https://ictinstitute.nl/the-artificial-intelligence-impact-assessment/> accessed 5 September 2019.

Any regulatory authority that assesses the potential for "accountability gaps" in existing legal and regulatory frameworks applicable to AI systems should adopt a balanced approach that encourages innovation while militating against the risk of significant individual or societal harm.

Most importantly, MITLA posits that Government should not grant distinct legal personality to AI systems, as doing so would undermine the fundamental principle that humans should ultimately remain accountable for the acts and omissions of AI systems.

Principle 3: AI - Transparency and Explainability

MITLA agrees that as a basic principle stemming from the human-centric approach, the use of AI systems should be transparent and that every AI-based decision can and should be explainable in humanly understandable terms, so as to allow for a valid review and/or appeal process.

This essentially reflects Article 5 of the European General Data Protection Regulation (hereinafter "GDPR") which provides that holders of personal data (controllers) shall process it "in a transparent manner," making transparency one of the core principles of the legislation. Similarly, any AI rules relating to transparency should as a minimum reflect the principles contained in Article 5 of GDPR. Reference should also be made to California's Bill 1001 where there is a requirement for the owner of an online chatbot to disclose, to the person interacting with it, that they are not communicating with a real human person.[28]

Furthermore, transparency in AI might also require the provision of information on the source and nature of the data used (the "lineage of data"),[29] notably by maintaining records of the origin of data.

---

[28] Senate Bill No. 1001 – Bots: Disclosure
<https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001>
accessed 5 September 2019.

[29] Singapore, Personal Data Protection Commission, "Discussion Paper on Artificial Intelligence (AI) and Personal Data—Fostering Responsible Development and Adoption of AI" (5 June 2018), p 7 <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf> accessed 5 September 2019.

With respect to "explainability" in AI, this refers to a duty to provide information about how exactly a certain output was produced. An explainable AI system, or XAI, is a system that provides explanations on its "thinking" process.[30] It implies, as stated by the European Commission's Draft Ethics guidelines for trustworthy AI that "AI systems [must] be auditable, comprehensible and intelligible by human beings at varying levels of comprehension and expertise."[31]

Transparency and explainability, taken together, should allow for a better monitoring of the lawfulness of the development, deployment and use of AI systems in general and should serve as an important building block for all the AI regulatory landscape. Like any other tool, policy or procedure used in a society, persons or institutions using AI systems are should comply with the applicable laws.

Rather than a one-size-fits-all approach, any proposal should factor in a gradual and context-sensitive approach, whereby the greater the impact of a decision or an action taken by an AI system on an individual or a group, the greater the obligation on the designer or user of the AI system to be transparent about its use and make the technology explainable to those affected. This should serve as a balancing test, taking its inspiration from various legal tests that seek to accommodate multiple parties' rights and preferences at one. In any case, even reasonable transparency and explainability rules will have to allow for a case-by-case assessment that may limit the principles when needed to protect legitimate third-party interests. The intensity of the obligations of transparency and explainability will depend on the context of the decision and its consequences for the person subject to it. Such intensity must therefore balance the interests of the person subject to the decision and the interests of the organisation making the decision.

---

[30] David Gunning, "Explainable Artificial Intelligence (XAI)" (Darpa, 2018) <https://www.darpa.mil/program/explainable-artificial-intelligence> accessed 5 September 2019.
[31] European Commission's High Level Expert Group on Artificial Intelligence, "Draft Ethics Guidelines for Trustworthy AI" (18 December 2018) p 10 <https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai> accessed 5 September 2019.

This will eventually assist in the preservation of the public's trust in AI systems thereby ensuring meaningful accountability of an AI system's developers, deployers and users, and to assist in the demonstrability of whether the decisions made by an AI system are fair and impartial.

Most importantly, AI systems should aim to be designed from the most fundamental level upwards to promote transparency and explainability by design and not as an afterthought, again reflecting the "privacy by design" principle found under GDPR.

Principle 4: AI – Fairness and Non-Discrimination

Like most human creations, AI systems tend to reflect the goals, knowledge and experience of their creators. The challenge to fairness arises when the systems are used to make decisions we would not have anticipated or the decisions are or appear to be biased, discriminatory or just plain "unfair."

With respect to policy considerations, MITLA believes that the issue of fairness in AI is not a straightforward one and will be complex and technically challenging.

It should be noted that the principles of equality and non-discrimination are part of the foundations of the rule of law where individuals are protected from discrimination not only by international treaties but also through national laws and regulations. Fairness and transparency should go hand in hand and should serve as a cornerstone for accountability. Improving the transparency of algorithms and training data is among the most important tasks when making use of AI systems, and is essential for preventing discrimination.

The EU's "Ethics Guidelines for Trustworthy AI"[32] recognise the importance of education and awareness to foster an ethical mind-set. In this sense, the guidelines stress that providing education regarding AI systems and their impact is crucial, and identify generating this awareness as a non-technical method to work towards fairer AI systems. This report aims to serve as a guide to the development of "Trustworthy AI," a concept which has two components: (1) its development, deployment and use complies with fundamental rights and

---

[32] Ibid.

applicable regulation as well as respects core principles and values, ensuring an ethical purpose; and (2) it is technically robust and reliable.

The document holds that ensuring an "ethical purpose" in AI requires a human-centric approach, founded in the principles of beneficence, non-maleficence, autonomy of humans, justice and explicability. Based on the necessity of safeguarding these principles, the HLEG on AI identifies a series of requirements for Trustworthy AI, which should be implemented from the earliest design phase. The concept of "Non-discrimination" is essential for ensuring the principle of justice, and therefore achieving Trustworthy AI.

Additionally and in line with MITLA's proposal regarding the need to create an intrinsic link between AI and fundamental human rights, the Toronto Declaration[33] tackles AI from the framework provided by human rights law. It focuses on the human rights to equality and non-discrimination, and the international framework that is in place to protect them. From the positive obligations which states have to protect human rights, the Toronto Declaration derives principles related to machine learning. It holds that states have a positive obligation to protect against discrimination by private sector actors and to promote equality, including through oversight and binding laws. Further, it states that these obligations also apply to public use of AI. The Declaration provides that in public sector systems, states must identify risks, ensure transparency and accountability, enforce oversight and promote equality. It should be clear therefore that decisions based on AI systems should be fair and non-discriminatory, judged against the same standards as decision-making processes conducted entirely by humans.

---

[33] --, 'The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems' *AccessNow.org* (RightsCon Toronto, 16 May 2018) https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/ accessed 5 September 2019.

The use of AI systems by organisations that develop, deploy or use AI systems, including government should not serve to exempt or attenuate the need for fairness, although it may mean refocusing applicable concepts, standards and rules to accommodate AI.

Any AI development should be designed to prioritise fairness. This would involve addressing algorithms and data bias from an early stage with a view to ensuring fairness and non-discrimination. Additionally, AI systems with an important social impact could require independent reviewing and testing on a periodic basis.

Principle 5: AI – Safety and Reliability

As stated by the EU HLEG on AI, safety is "about ensuring that the system will indeed do what it is supposed to do, without harming users (human physical integrity), resources or the environment. It includes minimizing unintended consequences and errors in the operation of the system." The importance for technologies that could affect life and well-being to have safety mechanisms in place that comply with existing norms, including legislation, can never be underrated. This is why when designing new technologies, it is important to keep a wide range of health and safety aspects in mind, including, e.g., ergonomics and mental stress.

With respect to reliability in the context of AI, the HLEG on AI underlined the requirement for algorithms to be secure, reliable, and robust enough to deal with errors during the phases of design, development, execution, deployment and use. Algorithms should also cope adequately with erroneous outcomes.

In the wider picture, applying European applicable safety and reliability laws and regulations to AI often proves difficult given how long ago these laws and regulations were put into force. Nevertheless, many European laws and regulations are technology neutral, which means it should be possible to apply these to AI use cases.

Government can play a major role in transition to AI systems, through various different forms of initiatives and interventions. For example, "command and control" regulation implies that the government threatens to impose sanctions aimed at bringing firms' interests into alignment with the interests of society

whilst meta-regulation and self-regulation are imposed by a non-governmental actor and are enforced by the regulated entity itself.

At this juncture one has to accept that effective regulation of algorithms might be difficult because AI cannot be defined easily.[34] Regulatory problems arise at the research and development stage (ex ante), and when the AI will be released into the world (ex post). In the ex post phase, there are problems with foreseeability and control. In the ex ante phase, problems with regard to diffuseness, discreteness and opacity can arise.

The diffuseness problem for example, suggests a need for global coordination in relation to AI regulation, which, MITLA believes, has to be kept at the forefront of any local discussion of Maltese AI regulation.

A system designed to autonomously make decisions will only be acceptable if it operates on the basis of clearly defined principles and within boundaries limiting its decision-making powers. Governments and organisations developing, deploying or using AI systems should validate the underpinning ethical and moral principles as defined periodically to ensure on-going accurateness. Furthermore, governments and organisations developing, deploying or using AI systems should recall that ethical and moral principles are not globally uniform but may be impacted e.g., by geographical, religious or social considerations and traditions. To be accepted, AI systems might have to be adjustable in order to meet the local standards in which they will be used.

MITLA posits that Government should require and organisations to test AI systems thoroughly to ensure that they reliably adhere, in operation, to the underpinning ethical and moral principles and have been trained with data which are curated and are as 'error-free' as practicable, given the circumstances.

Principle 6: AI – Open Data and Fair Competition

---

[34] John Danaher, "Is effective regulation of AI possible? Eight potential regulatory problems" (*PhilosophicDisquisitions*, 27 September 2018), <https://ieet.org/index.php/IEET2/more/Danaher20180927> accessed 5 September 2019.

Like any other new technology, the commercial development and deployment of AI-based solutions takes place within the standard legal frameworks for business activities. Competition (anti-trust) laws are a key part of this compliance challenge.

Whilst data is often considered to be proprietary and increasingly seen as a monetisable asset, control of data could potentially generate a market-distorting advantage in the development of AI. As the UK's Furman Report recently noted,

"to the degree that the next technological revolution centres around artificial intelligence and machine learning, then the companies most able to take advantage of it may well be the existing large companies because of the importance of data for the successful use of these tools."[35]

The impact of large datasets on competition has been considered in a number of fora, including the UK Furman Report, the OECD Roundtable on "Big Data: Bringing Competition Policy to the Digital Era,"[36] the joint report of the French and German competition authorities on "Competition Law and Data,"[37] and the report of the Competition Bureau Canada on "Big data and innovation: key themes for competition policy in Canada."[38]

MITLA believes that any local efforts towards creating an AI framework should adequately consider this aspect of AI.

---

[35] Report of the Digital Competition Expert Panel, 'Unlocking digital competition' <https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel> accessed 5 September 2019.

[36] OECD, 'Big data: Bringing competition policy to the digital era' <https://www.oecd.org/competition/big-data-bringing-competition-policy-to-the-digital-era.htm> accessed 5 September 2019.

[37] *Auterité de la concurrence*, 'Competition Law and Data' (10th May 2016) <http://www.autoritedelaconcurrence.fr/doc/reportcompetitionlawanddatafinal.pdf> accessed 5 September 2019.

[38] Report of the Government of Canada, 'Big data and innovation: key themes for competition policy in Canada' (19 February 2018) <https://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/eng/04342.html> accessed 5 September 2019.

There is wide acknowledgement of the importance of data as a driver of the digital economy but no consensus as to whether current competition regimes are sufficient to ensure effective competition in digital markets. Overall, the fact that there may be external or regulatory constraints on access to data is not treated as "unfair competition" or as a problem in itself in competition analysis, but as an aspect of the conditions of competition in a particular market.

EU Commissioner Margrethe Vestager has also expressed concerns about the control of data.[39] Given the importance of data to AI, the control of data also presents concerns about the ability to foster innovation when data is privately controlled by limited number of entities. Both the UK Government[40] and Commissioner Vestager, among others, have expressed such concerns.

The concerns around the risks from "data-opolies" and concentrations of data ownership have spurred calls for legal frameworks beyond competition law enforcement to address these issues. This requires examining options such as data trusts, data sharing agreements, and public sector programs to facilitate access to data.

One therefore must not allow the need for large quantities of data and/or the complexity of the technology to restrict the development and deployment of AI. There is increasing concern that businesses with huge datasets and an early lead in building AI applications could gain an advantage which is difficult to contest, with the commensurate concern that this could, in its worst cases, hinder innovation.

There is debate (and as yet no consensus) as to whether existing competition and antitrust legislative frameworks are sufficiently adaptable to offer strong deterrents and effective remedies to any market distortions which may emerge going forwards.

---

[39] Mackenzie Stuart Lecture, 'Making the data revolution work for us' (4 February 2019) <https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/making-data-revolution-work-us_en> accessed 5 September 2019.

[40] UK, Department for Business, Energy and Industrial Strategy, "Modernising Consumer Markets: Green Paper" (April 2018) https://www.gov.uk/government/consultations/consumer-green-paper-modernising-consumer-markets accessed 5 September 2019.

Open data and open source software are both important trends in this respect which should be encouraged and supported. Government should lead by example, where possible, in opening up datasets (with consents, pseudonymisation or aggregation etc., as necessary) and by investing in AI development resources, facilities and centres of excellence.

Given the expected strategic importance of AI, MITLA recommends that competition law enforcement authorities should proactively monitor the progress of this technology and the activities of those in the vanguard of its development as an administrative priority. International debate and discussion to develop best practice through fora such as the OECD or the International Competition Network should be strongly encouraged.

Government should undertake regular reviews to ensure that competition law frameworks and the enforcement tools available to the relevant enforcement authorities are sufficient and effective to ensure sufficient access to necessary inputs, and adequate choice, vibrant rivalry, creative innovation and high quality of output in the development and deployment of AI systems, to the ultimate benefit of consumers.

Government should foster and facilitate national infrastructures necessary to promote open access to datasets to all elements of society having a vested interest in access to such datasets for research and/or non-commercial use. In this regard, governments should give serious consideration to two-tier access models which would allow for free access for academic and research purposes and paid-for access for commercialised purposes.

Government should support open data initiatives in the public or private sector with guidance and research to share wide understanding of the advantages to be gained from open access data, the structures through which datasets can be shared and exchanged, and the processes by which data can be made suitable for open access (including API standardisation, pseudonymisation, aggregation or other curation, where necessary).

Government should ensure that the data held by public sector bodies are accessible and open, where possible and where this does not conflict with a

public sector mandate to recover taxpayer investment in the collection and cura-
tion of such data. Private sector bodies such as industry organisations and trade
associations should similarly support and promote open data within their
industry sector, making their own datasets open, where possible.

Any sharing or licensing of data should be to an extent which is reasonable in
the circumstances and should be in compliance with legal, regulatory,
contractual and any other obligations or requirements in relation to the data
concerned (including privacy, security, freedom of information and other
confidentiality considerations).

However, Government should at a minimum advocate accessibility through open
source or other similar licensing arrangements to those innovative AI systems
which may be of particular societal benefit or advance the "state of the art" in
the field via, for example, targeted incentive schemes.

Organisations that develop, deploy or use AI systems should design, develop
and deploy AI systems in a "compliance by design" manner which ensures
consistency with the overarching ethos of subsisting competition/antitrust
regimes to promote free and vibrant competition amongst corporate enterprises
to the ultimate benefit of consumers.

Principle 7: AI - Privacy

There is an inherent and developing conflict between the increasing use of AI
systems to manage private data, especially personal data; and the increasing
regulatory protection afforded internationally to personal and other private data.

Regulating the privacy implications of AI systems should be carried out in a
manner that acknowledges the specific characteristics of AI and that does not
unduly stifle AI innovation.
Furthermore, organisations that develop, deploy and use AI systems should
analyse their current processes to identify whether they need be updated or
amended in any way to ensure that the respect for privacy is a central
consideration.

AI systems create challenges specifically in relation to the practicalities of meeting of requirements under a number of national legislative regimes, such as in relation to consent and anonymization of data. Accordingly, organisations that develop, deploy or use AI systems and any national laws that regulate such use shall make provision for alternative lawful bases for the collection and processing of personal data by AI systems.

Moreover, organisations that develop, deploy or use AI systems should consider implementing operational safeguards to protect privacy such as privacy by design principles that are specifically tailored to the specific features of deployed AI systems.

Additionally, MITLA proposes that legislation should require that the development, deployment and use of AI systems should factor in the appointment of an AI Ethics Officer, in a role similar to Data Protection Officers under the GDPR, but with specific remit to consider the ethics and regulatory compliance of their use of AI.

Although there are challenges from a privacy perspective from the use of AI, in turn the advent of AI technologies could also be used to help organisations comply with privacy obligations.

Principle 8: AI – AI and Intellectual Property

It is generally accepted that there must be incentivisation and protection for innovation if it is to attract investment and be brought to the greater good of society. Within the scope of a national AI plan therefore, it must be made clear by Government that organisations must be allowed to protect rights in works resulting from the use of AI, whether AI-created works or AI enabled works. Typically such steps would include asserting or obtaining copyrights, obtaining patents, when applicable, and seeking contractual provisions to allow for protection as trade secrets and/or to allocate the rights appropriately between the parties.

MITLA however contends that at present, local, European and international IP laws are insufficiently equipped to deal with the creation of works by autonomous AI whilst at the same time care needs to be taken not to take steps

which will amount to overprotection, as this could prove detrimental to the ultimate goal of IP protection.

At the same time, one should be very cautious, at this early juncture in AI policy formulation, to embark into any revision of existing IP laws.

One should nevertheless explore the introduction of appropriate legislation (or the interpretation of existing laws) to clarify IP protection of AI-enabled as well as AI-created works, without seeking to create any new IP right at this stage. More importantly, when amending existing or implementing new IP laws, governments should seek adequately to balance the interests of all relevant stakeholders.

Finally on this topic, MITLA recommends that Government should also explore a consensus in relation to AI and IP rights to allow for unhindered data flows across borders and the rapid dissemination of new technologies and seek to address these issues through an international treaty.

Chapter 3: Industry and Trustworthy AI

The development of Malta's vision in ethical & trustworthy AI requires the participation of a large spate of stakeholders. Whilst studying emerging AI frameworks, and their ethical charters in other states, MITLA noted the critical participation of industry in the development and shaping of trustworthy AI.

Chapter 3 (Governance and Control Practices) of the Consultation Document[41] lays out a broad set of control practices to develop the unique trust conditions necessary for ethical AI. It is then argued that '*AI practitioners should use their judgement to understand their suitability and applicability*' and that ultimately the practitioners should observe domestic and international laws. Arguably, industry requires a clearer charter for adoption which is developed hand-in-hand with industry practitioners. Drawing experience from other industries (including the financial services sector in Malta which is presently undergoing a degree of reputational harm) it is critical to set out internal and external governance processes which are measurable, known, continuously improved and certifiable – and which provide pragmatic and directional insights.

The following areas highlight five (5) gaps in the Consultation Document which MITLA believes could be better addressed. The principles are based on research conducted by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems[42].

A) Data Agency

Digital consent is a misnomer in its current manifestation. Terms and conditions or privacy
policies are largely designed to provide legally accurate information regarding the usage of people's data to safeguard institutional and corporate interests, while often neglecting the needs of the people whose data they process. "Consent fatigue", the constant request for

---

[41] Malta.AI, Malta: Towards Trustworthy AI, August 2019.

[42] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. IEEE, 2019.

agreement to sets of long and unreadable data handling conditions, causes a majority of users to simply click and accept terms in order to access the services they wish to use. General obfuscation regarding privacy policies, and scenarios like the Cambridge Analytica scandal in 2018, demonstrate that even when individuals provide consent, the understanding of the value regarding their data and its safety is out of an individual's control. This existing model of data exchange must not erode human agency in the algorithmic age.

B) Effectiveness

Industry shall provide evidence of the effectiveness and fitness for purpose of the AI system deployed. The responsible adoption and deployment of AI solutions are essential if such systems are to realize their many potential benefits to the well-being of both individuals and societies. AI solutions will not be trusted unless they can be shown to be effective in use. Operators and other users will therefore benefit from measurement of the effectiveness of the AI system question. To be adequate, effective measurements need to be both valid and accurate, as well as meaningful and actionable. And such measurements must be accompanied by practical guidance on how to interpret and respond to them.

C) Competence

Industry shall specify and adhere to the knowledge and skill required for safe and effective operation of AI systems. AI systems can, and often do, make decisions that previously required human knowledge, expertise, and reason. Algorithms potentially can make even better decisions, by accessing more information, more quickly, and without the error, inconsistency, and bias that can plague human decision-making. As the use of algorithms becomes common and the decisions they make become more complex, however, the more normal and natural such decisions appear.

Operators of AI systems can become less likely to question and potentially less able to question the decisions that algorithms make. Operators will not necessarily know the sources, scale, accuracy, and uncertainty that are implicit in applications of AI systems. As the use expands, more systems will rely on machine learning where actions are not pre-programmed and that might not

leave a clear record of the steps that led the system to its current state. Standards for the operators are essential.

Operators should be able to understand how the AI system has reached its decisions, the information and logic on which the system relied, and the effects of those decisions. Creators of AI solutions should thus take an active role in ensuring that operators of their technologies have the knowledge, experience, and skill necessary not only to use the systems provided, but also to use such safely and appropriately, towards their intended ends.

D) Ethics in industry curricula

AI engineers and design teams do not always thoroughly explore the ethical considerations implicit in their technical work and design choices. Moreover, the overall science, technology, engineering, and mathematics (hereinafter "STEM") field struggles with the complexity of ethical considerations, which cannot be readily articulated and translated into the formal languages of mathematics and computer programming associated with algorithms and machine learning.

Ethical issues can easily be rendered invisible or inappropriately reduced and simplified in the context of technical practice. This problem is further compounded by the fact that many STEM programs do not sufficiently integrate applied ethics throughout their curricula. When they do, often ethics is relegated to a stand-alone course or module that gives students little or no direct experience in ethical decision-making and fosters the impression that it is not an essential component of development. Ethics education should be meaningful, applicable, and incorporate best practices from the broader field.

Ethics training needs to be a core subject for all those in the STEM field, beginning at the earliest appropriate level and for all advanced degrees.

E) Interdisciplinary collaborations

More institutional resources and incentive structures are necessary to bring AI engineers
and designers into sustained and constructive contact with ethicists, legal scholars, and social scientists, both in academia and industry. This contact is

necessary as it can enable meaningful interdisciplinary collaboration and shape the future of technological innovation. More could be done to develop methods, shared knowledge, and lexicons that would facilitate such collaboration.

F) Equal Availability

Vastly different power structures among and within countries create the risk that AI deployment accelerates, rather than reduces, inequality in the pursuit of a sustainable future. It is imperative that all humans, in any condition around the world, are considered in the general development and application of AI systems to avoid the risk of bias, excessive inequality, classism, and general rejection of these technologies. With much of the financial and technical resources for AI development and deployment residing in High-Income Countries (hereinafter "HIC"), not only are AI benefits more difficult to access for Low and Middle-income countries (hereinafter "LMIC") populations, but those AI applications that are deployed outside of HIC realities may not be appropriate. This is for reasons of cultural/ethnic bias, language difficulties, or simply an inability to adapt to local internet infrastructure constraints.

If AI capacity and governance problems, such as relevant laws, policies, regulations, and anticorruption safeguards, are addressed, LMIC could have the ability to use AI to transform their economies and leapfrog into a new era of inclusive growth. Indeed, AI itself can contribute to good governance when applied to the detection of corruption in state and banking institutions, one of the most serious recognized constraints to investment in LMIC.

Chapter 4: Comparative analysis – how Malta's efforts compare to those of other states

While undertaking its analysis of the Consultation Document and formulating its response to same, MITLA considered that benefits could be reaped from studying the manner in which other countries, including other member states, have so far tackled the ethical underpinnings of their respective national strategies. Having an extensive and detailed ethics plan, the United Kingdom was considered as being the ideal candidate for comparative purposes.

The UK Government's industrial strategy, similarly to Malta's, plans on putting the UK at the forefront of the AI and data revolution[43]. The same strategy identified the need of ethical use of data and AI that gives confidence and clarity to citizens and business. To this end, the UK has invested over seven (7) million pounds in a new Centre for Data Ethics and Innovation (hereinafter "CDEI"): an advisory body intended to review the existing governance landscape and advise the UK government on how it can enable and ensure ethical, safe and innovative uses of data, including AI[44].

In the November 2018 consultation surrounding the set-up of the CDEI, the UK Government proposed areas in which the CDEI could undertake projects to strengthen the governance of data and AI uses. Within these areas, the CDEI is encouraged to identify specific policy or regulatory actions required to address or prevent barriers to innovative and ethical uses of data, and on which the CDEI could publish clear and evidence-based recommendations to Government. These areas in many ways echo the ITECHLAW principles tackled extensively in chapter 2 of this reaction document and consist of:

---

[43] HM Government, 'Industrial Strategy: Building a Britain fit for the future' (November 2017) p14
<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730048/industrial-strategy-white-paper-web-ready-a4-version.pdf?_ga=2.57742307.853808025.1566907171-430880583.1566761608> accessed 5 September 2019.

[44] Ibid, p 40.

Targeting – The CDEI is encouraged to address the manner in which data and AI can produce insights about the public or users' behaviour and emotions. This should not have the effect of limiting the information and choices available to users and to the public, and should not be used to influence, manipulate, or control behaviour in harmful ways.

Fairness and Bias – Data should promote fair and equitable outcomes, minimising unconscious biases.

Transparency – Decisions and recommendations made by data technologies should be easily interpretable or explainable. The CDEI was invited to delve into the extent to which decisions must be explained in the different contexts, as well as the level of retention of human control over decision-making.

Liability – Autonomous systems naturally challenge traditional notions of accountability and the CDEI should look into developing potentially new models of liability to reflect this challenge.

Data Access – Data creation, sharing and trading is at the core of any leading digital economy and government, and requires sufficient data sharing frameworks.

IP and Ownership – Establishing the owner of IP, and of datasets, training data, source code among others, is necessary in order to ensure that innovation and creativity remains rewarded, while at the same time ensuring that the persons responsible for the legitimate processing of personal data remain identified at all times[45].

---

[45] Department for Digital, Culture, Media & Sport, 'Consultation outcome: Centre for Data Ethics and Innovation Consultation' (updated 20 November 2018), para 3.6 <https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation> accessed 5 September 2019.

The CDEI also has a mandate to test and supplement as necessary existing data protection legislation, as technology, business models and societal views on data and AI continue to develop[46].

The CDEI 2019/2020 work programme confirmed the above list by verifying that it will be reviewing the targeting nature of AI, as well as the Bias factor[47], i.e. the first two on the list of areas identified by the UK Government within the consultation document. In July 2019, the CDEI published the Interim report Review into online targeting. The methodology of the review tackled three aspects: the first was determining whether the use of the technology aligns with public values, the second was an analysis of whether current regulatory mechanisms succeed in delivering their intended outcomes, and the third was the proposal of solutions relating to any technical, legal or other mechanisms that ensure that the use of online targeting is consistent with the law and public values[48].

MITLA commends the model adopted by the UK in light of the fact that this Centre is specifically focused on ethical use of data and AI, and is not equally tasked with the fulfilment of the entire AI Strategy proposed by the UK Government. This model ensures that the ethical use of data is given the priority it deserves, while ensuring that a number of individuals are in continuous contact with the industry, the general public and academics to propose to identify areas that require legislative action, and proposing relevant solutions to Government. Within the Consultation Document, Malta's AI taskforce has already identified that there can be no solution that fits all scenarios, however, breaking down the

---

[46] Ibid, para 3.7.

[47] Centre for Data Ethics and Innovation, 'Introduction to the Centre for Data Ethics and Innovation'
p 7, 8
<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf?_ga=2.100220503.853808025.1566907171-430880583.1566761608> accessed 5 September 2019.

[48] Centre for Data Ethics and Innovation, 'Interim report: Review into online targeting' (July 2019)
<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819169/Interim_report_-_review_into_online_targeting.pdf?_ga=2.136246904.853808025.1566907171-430880583.1566761608> accessed 5 September 2019.

different aspects of AI use and addressing the ethical implications of each, could lead to more precise considerations and ensuing recommendations.

A further consideration is that, as already pointed out in relation to the ITECHLAW principles touched upon in chapter 2, Malta did not make any reference to IP and ownership of data. No reference has been made within the Consultation Document as to the analysis of the impact of AI on traditional IP concepts, for example, the notion of the 'inventor' in AI-created inventions[49].

The Taskforce is encouraged to look into whether or not AI-generated inventions and creations merit IP status, and the ethical implications of such considerations. The ownership of the data itself is also a considerable consideration in this respect: should data be subject to proprietary categories, or should it be freely available to feed into the development of AI?

On a final note regarding the UK approach and in contrast with same, the Consultation Document only makes a very brief reference to the issue of liability; it notes that risk assessment of AI systems causing harm or damage to users or third parties should include a consideration of liability and consumer protection rules. By virtue of this feedback, MITLA calls for a more detailed focus on the implications that AI technology has on the current legal framework tackling liability.

On another note, MITLA notes developments in other states, including the recent announcement of the collaboration between France and Germany to invest public money in data storage facilities in a combined effort to spur investment in AI and address Europe's current dependence on data centres located abroad, including in the US and China, citing as one of the main reasons for this move the protection of European data and the creation of sovereign data structures.[50]

---

[49] WIPO, 'Artificial intelligence and intellectual property: an interview with Francis Gurry' (September 2018) <https://www.wipo.int/wipo_magazine/en/2018/05/article_0001.html> accessed 5 September 2019.

[50] --, 'France, Germany to propose public investment in data centers for AI' (28 August 2019)

MITLA believes that such an effort shows the willingness of certain European member states to put Europe back at the forefront of AI development. MITLA hopes that Malta will not shy away from this European call for action and will participate by investing its finances and resources into this cause, particularly considering Malta's economic reliance in data-driven industries.

<https://kfgo.com/news/articles/2019/aug/28/france-germany-to-propose-public-investment-in-data-centers-for-ai/931785/> accessed 5 September 2019.

Concluding remarks

Despite the tight deadlines imposed for the submission of feedback, MITLA commends the public consultation launched on building an ethical AI framework for Malta, being that such efforts go hand in hand with the principle of building AI systems which are human-centric and which reflect, echo and promote the interests of the societies in which they are implemented.

MITLA hopes and recommends that the AI Taskforce follows up on the feedback provided and does not engage in any knee-jerk reactions to industry developments which are geared solely towards attracting foreign investment, before having first taken on board and implemented the suggestions made to bolster Malta's legal framework in line with industry requirements and discussions undertaken at EU level and internationally.

MITLA finally encourages the Taskforce to involve all stakeholders and publish the results of this public consultation in a timely manner so as not to lose the momentum gathered by this public initiative towards the formulation of a Maltese ethical framework for the development of AI.[51]

---

[51] This document has been prepared by: Dr Antonio Ghio, Dr Gege Gatt, Dr Lena Sammut, Dr Deo Falzon and Dr Yanica Sant, board members of MITLA, 2019.

The Malta IT Law Association (MITLA) is active in the research, discussion and circulation of information on legal developments taking place on the international plane and within the European Union with respect to ICT Law and the knowledge economy.

MITLA is registered as a Voluntary Organisation (VO/1166) in terms of Article 3 of the Voluntary Organisations Act 2007 (Act No, XXII of 2007), Malta.

www.mitla.org.mt
info@mitla.org.mt