Regulating Artificial Intelligence: A Technology Regulator's Perspective

Joshua Ellul
joshua.ellul@um.edu.mt
Malta Digital Innovation Authority & University of Malta
Malta

Gordon Pace
gordon.pace@um.edu.mt
Department of Computer Science, University of Malta
Malta

Stephen McCarthy
stephen.mccarthy@mdia.gov.mt
Malta Digital Innovation Authority
Malta

Trevor Sammut
trevor.sammut@mdia.gov.mt
Malta Digital Innovation Authority
Malta

Juanita Brockdorff
JuanitaBrockdorff@kpmg.com.mt
KPMG
Malta

Matthew Scerri
MatthewScerri@kpmg.com.mt
KPMG
Malta

ABSTRACT

Artificial Intelligence (AI) and the regulation thereof is a topic that is increasingly being discussed and various proposals have been made in literature for defining regulatory bodies and/or related regulation. In this paper, we present a pragmatic approach for providing a technology assurance regulatory framework. To the best of our knowledge, this work presents the first national AI technology assurance legal and regulatory framework that has been implemented by a national authority empowered through law to do so. Aiming to both provide assurances where required and not stifling innovation yet supporting it, it is proposed that such regulation is not to be mandated for all AI-based systems but rather should provide a voluntary framework and only be mandated in sectors and activities as deemed necessary by other authorities or laws for regulated and critical areas.

CCS CONCEPTS

• **Social and professional topics → Governmental regulations**.

KEYWORDS

artificial intelligence, regulation

ACM Reference Format:
Joshua Ellul, Gordon Pace, Stephen McCarthy, Trevor Sammut, Juanita Brockdorff, and Matthew Scerri. 2021. Regulating Artificial Intelligence: A Technology Regulator's Perspective. In *Eighteenth International Conference for Artificial Intelligence and Law (ICAIL'21), June 21–25, 2021, São Paulo, Brazil.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3462757.3466093

1 INTRODUCTION

Issues concerning the design of legal and regulatory frameworks for Artificial Intelligence (AI) have been a topic of discussion and debate for the past few decades. Much of the debate inherits from discussions on regarding how to regulate technology and the regulation of computer systems, but reaches further due to the very nature of the potential for AI. In fact, one can argue that a substantial portion of the debate is due to this very potential, which brings together ethical issues, rights, perils and other aspects. Regardless of which school of thought one subscribes to, in a spectrum ranging from the requirement of generic principles [1], to specific laws[1], to advocating that regulation of such technology should be avoided, and focus should be on safety mechanisms [16], when the technology is used for applications that can directly or indirectly impact society then sufficient regulation (whether through law or otherwise) should be investigated (whether applied directly or indirectly to the technology). At the same time whilst some argue for mandatory regulation, many warn that regulation could stifle innovation [11].

In this paper, we do not purport to present a contribution to this philosophical debate, but rather our aim is a more pragmatic one — that of outlining and explaining the rationale behind a legal and regulatory framework addressing AI systems adopted by Malta. Whilst other regulatory frameworks and bodies have been proposed in literature (discussed in Section 5), it is to the best knowledge of the authors that the framework being presented herein is the first AI technology assurance legal and regulatory framework that has been implemented by a national authority (the Malta Digital Innovation Authority[2]) empowered through law [10] to do so.

Towards the end of the 2010's Malta built a framework for addressing the regulation of Innovative Technology Arrangements [9], in order to ensure better end user protection through the adoption of appropriate due diligence on the underlying technologies. Initially focusing on Blockchain, Smart Contracts and other Distributed Ledger Technologies (DLTs) [6], the legislation has since been extended to cover critical systems (through a legal notice[3]) and regulatory guidelines have been issued by the Malta Digital Innovation Authority (MDIA) for the regulation of arrangements which use an element of AI.

The aim of the paper is to provide a review of the regulatory framework proposed and to put it in the context of the ongoing AI regulation debate. One of the primary observations is that the need

ICAIL'21, June 21–25, 2021, São Paulo, Brazil

ACM ISBN 978-1-4503-8526-8/21/06.
https://doi.org/10.1145/3462757.3466093

[1] https://www.europarl.europa.eu/doceo/document/E-9-2019-002411_EN.html
[2] https://mdia.gov.mt
[3] https://legislation.mt/eli/ln/2020/389/eng/pdf

for precise definitions and objective measures in the legal framework, meant that the Maltese regulatory approach is founded on practical and auditable aspects, and is intended to address concerns with existing technology (as opposed to attempting to address possible issues arising from future development of AI technology, for example Artificial General Intelligence). To implement an AI regulatory framework intended for modern day technology and also in aim of not stifling innovation, the framework is primarily voluntary however may be mandated based upon the sector and/or risk associated with the activity within which the AI system is used or as deemed necessary by another lead authority or governing legislation. This sets the tone of much of the paper, but it is naturally endemic to any discussion of practical implementations of the regulation of technologies. The fast evolving nature of technology requires law-makers to address existing technology in a sound manner, but also in a way that is expected to be future-proof. A full version of this paper can be found in [7].

2 THE CASE FOR AI ASSURANCES

We start by highlighting issues related to AI-based systems which could result in systems operating incorrectly in relation to its intended functionality and thereafter build the case for instilling assurances. We concentrate on Artificial Narrow Intelligence (ANI) given that the state-of-the-art has not yet reached levels of Artificial General Intelligence (AGI) [3]. We will use the term AI throughout the rest of the paper to refer to AI that exists today — ANI.

Since the inception of software development, the fact that such systems occasionally fail has been accepted to be the norm. Although much work has gone into developing techniques to reduce the frequency and severity of such occurrences, we continue to experience software malfunction on a daily basis. The impact of such failure is contained as long as the software functions in a closed system i.e. it has no direct impact on the real world, but frequently software affects the real world in a direct or indirect manner. One finds reports of many catastrophic failures in literature and newspaper reports, with effects ranging from huge financial losses to critical infrastructure failure and even loss of human life. AI systems are no exception when it comes to incorrect behaviour and even when the algorithms themselves are correctly implemented, incorrect behaviour might emerge. For instance, a correct implementation of a machine learning based algorithm may still learn wrong due behaviour due to insufficient training, biases and unbalances that may exist within datasets, etc.

Undoubtedly AI systems should undergo standard quality assurances processes, not only for functional correctness of the algorithms themselves but also with respect to the behaviour emergent following training. However, testing of AI systems is only as good as the coverage of training data, iterations and permutations and use cases which are undertaken. Once an AI system is deployed and it encounters an event that it was not trained to handle it may well end up handling it incorrectly. More so, if it is continuously learning in a live environment it may be exposed to certain situations which could affect its behaviour negatively.

Part of the challenge is that many AI-based techniques function as black-boxesfor which reason one finds extensive research towards explainable AI. The past decade has seen various infamous cases where unexpected behaviour emerged from AI systems, sometimes of a controversial or even safety-critical nature. The increasing concern is not only to do with cases that have emerged but also based on the reality that more and more systems are becoming computerised and automated. One often-referenced cliché is that of automated scoring systems [13] in which discrimination is unacceptable, highlighting the need to ensure bias in datasets is removed and attempts made to remove discriminatory features during training.

The concerns highlighted above demonstrate the need to ensure that sufficient assurances are put in place to ensure that AI algorithms are implemented correctly, and that their behaviour is as expected and does not introduce any unwanted biases. Indeed, many are advocating for such regulatory frameworks to be developed and applied to AI systems, but it whether such frameworks should be mandatory for all AI-based systems is debatable. We now follow with a case for why such frameworks should not always be mandated, and that they should not be focused on the technology but should be focused on the sector or activity that the AI is being used within/for.

3 THE CASE FOR VOLUNTARY ASSURANCES OF AI AND MANDATORY ASSURANCES OF REGULATED AND CRITICAL ACTIVITIES

Setting aside AGI, when it comes to ANI should such frameworks always be mandated? The same AI framework, for instance identifying user preferences, can be the engine behind a wide range of applications, from a personal movie recommendation system, to a social network targeted advertising campaign to influence users in an upcoming election. The underlying infrastructure is application agnostic, but should such an underlying infrastructure be required to be regulated? More so, what difference does it make if an algorithm is AI-based or not and yet can be used for the same activity? Then, should we be talking about AI regulation at all? Or should we be focusing on software — or rather, the activity it is used for irrespective of how it is implemented?

Regulating all forms of AI would result in shackling and stifling innovation [11]. The definition of AI itself is controversial, and even if a definition is chosen, is it going to be clear what software is AI and what software is not? There are some algorithms which we can ascertain are universally accepted as AI, and some systems which are universally considered to not have aspects of AI within them, however what should be done about the rest? Could this approach not only stifle AI-innovation, but also other software based innovation?

Looking back at the principles of regulation though, we need to ask ourselves why is regulation of AI being proposed? Is it only because of end-of-the-world scenarios being painted which require AGI, which the state-of-the-art is currently not capable of? If so, then perhaps we should differentiate between any regulatory requirements for AGI and ANI. We propose that this should be done, at least in the interim until AGI is deemed to be upon us. We leave considerations for AGI as future work, and here will continue discussing aspects pertaining to ANI.

If AI is regulated even when applied to unregulated and non-critical activities, given the line between AI and software in general is blurred, and given that non-AI based techniques processes may

yield the same sort of undesirable outcomes, then why should not all software be regulated? We propose that mandatory regulation should be sector/activity-based and not on technology.

The question of what constitutes high-risk or defines whether a sector or activity should mandate this framework arises. This is to be left up to other lead authorities and laws of the land to decide. For financial affairs, a financial services authority (a separate body) may impose when a sector or activity should be mandated to undertake a technology audit (as proposed herein), or even if any levels of enhanced due diligence is required. Therefore, based on the above we make the argument that mandatory regulatory frameworks should not be technology-specific (or AI-specific), yet should be activity or sector-specific as defined and required per activity/sector.

AI technology-based assurances may not only be required for regulated activities, however various AI-based products and services may see benefit in providing assurances to various stakeholders. Therefore, the regulatory approach enables for technology-based assurances to also be offered on a voluntary basis (besides being mandated from lead authorities of respective sectors/activities).

Now, we present the AI technology assurance framework implemented by the Malta Digital Innovation Authority[4] which offers certification of AI systems on a voluntary basis where sought, or on a mandatory basis where other lead authorities or laws require it.

4 AN AI TECHNOLOGY ASSURANCE REGULATORY FRAMEWORK

We now present the AI Innovative Technology Arrangement (AI-ITA) regulatory technology assurance framework. Approaches for providing software assurances will invariably have a degree of commonality irrespective of the technology domain and also application domain within which the solution is categorised under. As such, this framework builds on the Innovative Technology Arrangement (ITA) [6] regulatory assurance framework overseen by the Malta Digital Innovation Authority (MDIA). Rather than mandating compliance and certification of all AI based systems, the regulatory framework is a voluntary one — unless a lead authority deems that such technology assurances are required. It is in this manner we believe innovation can still flourish, by only requiring mandatory oversight of sectors and activities that should require such oversight.

AI Innovative Technology Arrangement. The challenge with Artificial Intelligence ITAs (AI-ITAs), primarily revolves around identifying what constitutes AI. Rather than define what is an AI-ITA as a hard and fast rule, the guidelines take the approach of defining qualities and criteria that qualify software as an AI-ITA: (a) the ability to use knowledge acquired in a flexible manner in order to perform specific tasks and/or reach specific goals; (b) evolution, adaptation and/or production of results based on interpreting and processing data; (c) systems logic based on the process of knowledge acquisition, learning, reasoning, problem solving, and/or planning; (d) prediction forecast and/or approximation of results for inputs that were not previously encountered.

The above ensures that techniques and algorithms commonly associated with the wider AI field are captured and include anything from Deep Learning to Natural Language Processing and Optimisation Algorithms. The MDIA will also continue to monitor developments and update guidelines as required to include (and potentially exclude) defining features of what is/not classifed as an AI-ITA.

System Audits and Subject Matter Experts. The framework provides a structure for the Authority and applicant to work with independent (and approved) system auditors to be able to scrutinise to a fairly high level of detail the software itself as well as the manner with which it is being operated under the ISAE 3000 [12] standard for assurance. The audit of the software system itself is primarily conducted via a code review, whose aim is to ensure that the manner with which the AI-ITA is implemented accurately reflects what the organisation behind the AI-ITA are claiming in their technology blueprints. The rationale behind this is to ensure that any claims being done are truly reflected in the code, which enables the general public, who may not know what AI really is to gain trust in the system given that it stood up to scrutiny prior to the certificate being issued. Beyond the software, the certification mandates depending on the type of audit being undertaken and associated controls, to also give the general public assurances that the AI-ITA creator and operator are running the organisation in a manner that meets the standards set out by the MDIA. The certification therefore enables the general public to trust the creator, in the manner they build, maintain and run the AI system. Two main types of audits are required throughout an AI-ITA's lifetime: (i) first a 'Type 1 Systems Audit' is required which focuses on providing assurances with respect to functional correctness typically undertaken as an AI-ITA's first audit; and (ii) a 'Type 2 Systems Audit' which focuses on renewing previous assurances provided through a previous audit which factors in live data and operations associated with the system to assure the system assurances are still in place within the period under audit.

The audit process begins with the applicant submitting a request (in the form of an application) to the Authority, upon which the Authority will assess the applicant by reviewing the provided documentation around the AI-ITA and conduct its due diligence. Following this, the MDIA issues a Letter of Intent upon which the applicant will be able to appoint an MDIA approved Systems Auditor, and notify the MDIA of the appointment, for the MDIA to verify that the Systems Auditor has the required competencies (which the Authority has tested the system auditor for). The Systems Auditor will then conduct the audit as per the Authority's guidelines[5] and compile a report with their findings, which is issued to the MDIA for a review and a subsequent decision on whether the certificate is to be issued. Once issued, a further follow up audit must be conducted every time there is a material change in the AI-ITA (and on renewal after every two years).

Systems Audits are an integral part of the certification process as they provide the MDIA with an independent report on the particularities of the AI-ITA, specifically the code (and data) and whether it accurately reflects what is being disclosed in the blueprint, and the ongoing operations of the AI-ITA. Systems Audits are conducted by Systems Auditors, who must be independent from the AI-ITA and its operator, that are subject to approval by the Authority, and who need to meet a set of requirements (defined in the Systems Auditor guidelines) through their combined complement of Subject Matter

[4] https://mdia.gov.mt

[5] https://mdia.gov.mt/wp-content/uploads/2019/10/AI-ITA-Guidelines-03OCT19.pdf

Experts (SMEs) in the fields of IT audit, cybersecurity and technology with specialisation, in this case, in AI. The SMEs will be the primary individuals responsible for conducting systems audits, and must adhere to a set of requirements, such as ensuring that they meet a level of continuous professional education in the AI field[6].

This section describes the requirements that an AI-ITA must meet in order to qualify for certification.

ITA Blueprint. The Blueprint document is an essential document in the certification process as it is meant to provide a detailed description to the Authority on what the system does, how it is designed, and operated. Other than allowing the MDIA to evaluate whether AI-ITA certification is applicable, it is further intended to be used by the Systems Auditors as the document against which aspects such as the code is reviewed against. The blueprint also defines a minimum set of disclosures that must be disclosed to direct users (in English) in a non-technical manner, to be able to communicate the features and functionalities of the system and how it respects the ethical AI framework[7], limitations to prevention of bias, and the expected accuracy of the AI-ITA.

In a general (AI agnostic) sense, the detailed description must cover the functional capabilities of the AI-ITA, how the system is to be verified and tested to ensure the results meet expectations and what the operational limitations of the systems are. More specifically, for an AI solution the blueprint must include a disclosure of the AI techniques used and to justify why certification is being sought, and how specific risks are being managed and mitigated e.g. what is being done to ensure that the underlying dataset is unbiased. In a broader sense, the Blueprint must highlight the safety mechanisms in place and alignment with Malta's Ethical AI Framework.

ITA Harness. A crucial element that the AI-ITA framework proposes, and which needs to be highlighted clearly in the Blueprint is the ITA Harness. The ITA Harness provides a safety net for the process by monitoring activity inputs and outputs to ensure that the boundaries (which must also be disclosed in the Blueprint) are respected. Furthermore, the ITA harness must also be able to handle any anomalies it detects (such as outputs outside expected boundaries) in a manner which is also disclosed. The AI-ITA harness must also communicate with the Forensic Node (discussed next) to ensure that any anomalies are appropriately logged and can be investigated and rectified. While the harness may not apply to all AI-ITAs, the Authority requires that when it does not apply it must be justified adequately in the blueprint and accompanied with alternative plans of how the behaviour of the AI-ITA will be monitored and contained[8].

Forensic Node. The Forensic Node is another requirement mandated by the MDIA, and whose implementation and operation is also subject to the audit. The purpose of the Forensic Node is to *"store all relevant information on the runtime behaviour of the AI-ITA in real-time such as recording of inputs and outputs, and supporting data related to potential explainability of how an output was derived from a given input wherever applicable"*. This means that any inputs, outputs as well as data that supports how the system achieved the results it did must be stored in a secure data store in real-time. This highlights that the Forensic Node is not only used to support the assessment of (some of) the operating effectiveness of the controls during an audit, but may also be used to support legal compliance by the MDIA (or other authorities) and also enables a further layer of monitoring to be done (manually or automated) by a Technical Administrator (discussed next). It is important to note that the Forensic Node must be separate from the ITA Harness, in that the Forensic Node is more concerned with Data Logging, as opposed to the monitoring in relation to boundaries.

Technical Administrator. A Technical Administrator, a form of a service provider appointed by the AI-ITA to act as the final safeguard for the system, must be appointed and in place at all times. The Technical Administrator must be able to intervene, if required to do so by the MDIA, another authority or legally (such as in the event of a breach of law by the AI-ITA), to limit further impact to the users and where necessary limit or reverse losses. For example, consider an AI system that utilises reinforced learning and which, after a period of time, starts to exhibit discriminatory bias that goes against the principles laid down in the ethical AI framework and/or against the requirements of any laws or rules it must abide by. In this case the Technical Administrator must be able to halt the operation of the system to prevent further damage and revert to an older model (as may be mandated by a legal judgement). As such, this also imposes an indirect requirement for the AI-ITA to provide mechanisms to enable the Technical Administrator to conduct their actions as may be necessary (e.g. by ensuring regular snapshots of the machine learning models are kept to revert back to).

English Description and Consumer Protection. The system being certified is checked by the systems auditors who, amongst other things, ensure that its functionality matches that described in the blueprint in human-readable form (in English). If, post-deployment, the system exhibits behaviour contrary to this description against which it was certified, the Innovative Technology Arrangements and Services Act specifies that the English version prevails legally.

Auditing of Design and Development Processes. Systems Audits include oversight of the design and development process of the system-under-audit. Not only does such oversight cover traditional software engineering principles, but for systems including an element of AI also includes assurances that certain foundational principles have been taken into consideration in the process.

Build on a human-centric approach. The systems auditors ensure that the AI system was designed in a manner to support and assist humans without overriding the user into taking any unwanted decisions and the manner with which it operates musts be equitable and inclusive across different segments of society.

Adherence to applicable laws and regulations. It is crucial that behaviour induced by the system, including parts driven by AI, will be designed in a manner that adheres to the law.

Maximise benefits of AI systems while preventing and minimising their risks. It is crucial that any risks induced through the use of AI are identified and mitigated accordingly, including the setting up of controls to ensure fairness, transparency and resiliency to new AI-specific attack vectors.

Aligned with emerging international standards and norms around AI ethics. As the world is increasingly becoming globalised through technology, and which may be further amplified through the proliferation of AI systems, this objective was laid down to ensure that

[6] https://mdia.gov.mt/sa-guidelines/

[7] https://malta.ai/wp-content/uploads/2019/08/Malta_Towards_Ethical_and_Trustworthy_AI.pdf

[8] https://mdia.gov.mt/wp-content/uploads/2019/10/AI-ITA-Blueprint-Guidelines-03OCT19.pdf

Malta's ethical framework is aligned with similar ethical guidelines by the EU commission[9] and OECD[10].

The framework further builds on these principles by delineating a number of principles (such as Human Autonomy, Fairness, Prevention of Harm and Explicability) and proposes 63 controls of how these can be tested. While not all of these controls apply to all AI-ITAs, the AI-ITA must show that it has taken them into consideration and justify in the Blueprint (and ultimately top the users) those controls which do not apply.

5 RELATED WORK

The work presented herein is complementary and orthogonal to a number of different areas which we will now provide an overview.

Regulatory Bodies. The European Parliament had proposed for the setting up of an European Agency for robotics and artificial intelligence address technical, ethical and regulatory aspects [5], mostly driven from the need for transparency of automated systems handling personal data and the often impossibility of doing so due to trade secret protection [15]. A solution proposed was to allow for a trusted third party to undertake an audit of the system in question. On similar lines, to avoid differing domestic approaches, the need for an International Artificial Intelligence Organisation was highlighted [8]. Indeed, this would be a step in the right direction, however it is the opinion of the authors that the need for providing regulation should not wait for such an organisation to emerge, yet national authorities (such as the MDIA) could work together towards harmonisation and adapt to eventual international standards and guidance as it emerges.

International Standards. Whilst global software regulatory bodies do not (yet) exist, global standards do. The International Organisation for Standardization (ISO) has developed a number of different standards for use within the software domain[11]. Whilst, such software focused standards can be useful for global recognition within the framework described herein local national standards were required to be developed for the following reasons: (i) standards available to date do not provide guidelines or comprehensive control objectives specific for the artificial intelligence domain; (ii) mechanisms and roles for ensuring continuous monitoring and intervention are not defined [2, 4];and (iii) the authority is ultimately responsible and empowered through the MDIA Act to ensure audit integrity and quality whilst at the same time able to propose changes to legislation and guidelines. Once international standards adequate to adopt are developed national guidelines may be updated to make use of them (if deemed to meet the national requirements). That said, the authority has adopted and requires that audits are undertaken following the ISAE 3000 [12] standard which specifies generic (i.e. not software nor AI related) principles for quality management, ethical behaviour and performance for use in non-financial areas.

Other Non-technology Assurance Related Aspects. There is a large body of work looking at regulating the application of technology which is orthogonal to the to approach presented in this paper, including how to handle issues of liability, intellectual property and copyright, sector specific regulation (e.g. health, autonomous vehicles, finance, etc.), privacy and data protection, fundamental rights, profiling and anti-discrimination issues, competition law, and legal personality. Quite a number of ethical frameworks have been proposed— it suffices to note that *"at least 63 public-private initiatives have produced statements describing high-level principles, values and other tenets to guide the ethical development, deployment and governance of AI"* [14]. Within the framework proposed herein an ethical framework is referenced to, however the scope of such a discussion would warrant a paper of its own.

[9]https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[10]https://www.oecd.org/going-digital/ai/principles/

[11]https://www.iso.org/ics/35.080/x/

6 CONCLUSIONS

Moving towards a more efficient digital world has its benefits, however it brings various risks that need to be mitigated through adequate levels of technology assurances. In this paper we have highlighted the need for an AI technology assurance regulatory framework which is implemented in a manner that both promotes technology and does not stifle innovation, yet at the same time enforces assurances where required, and we have presented an implementation of a national AI regulatory framework that is overseen by a technology regulator. The guiding principle was that mandatory regulation of AI should be avoided, but other national regulators (e.g. finance, health, communications, etc.) can then work together with the technology-centric regulator to identify whether mandatory assurances are required.

REFERENCES

[1] Isaac Asimov. 1950. *I, robot.* Gnome Press.

[2] Olivier Boiral. 2011. Managing with ISO systems: lessons from practice. *Long Range Planning* 44, 3 (2011), 197–220.

[3] Nick Bostrom and Eliezer Yudkowsky. 2014. The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence* 1 (2014), 316–334.

[4] Francois Coallier. 1994. How ISO 9001 fits into the software world. *IEEE Software* 11, 1 (1994), 98–100.

[5] M Delvaux. 2016. *Motion for a European Parliament resolution: with recommendations to the commission on civil law rules on robotics.* Technical Report. Technical Report (2015/2103 (INL)), European Commission.

[6] Joshua Ellul, Jonathan Galea, Max Ganado, Stephen Mccarthy, and Gordon J Pace. 2020. Regulating Blockchain, DLT and Smart Contracts: a technology regulator's perspective. In *ERA Forum.* Springer, 1–12.

[7] Joshua Ellul, Stephen McCarthy, Trevor Sammut, Juanita Brockdorff, Matthew Scerri, and Gordon J. Pace. 2021. A Pragmatic Approach to Regulating Artificial Intelligence: A Technology Regulator's Perspective. *CoRR* abs/2105.06267 (2021). arXiv:2105.06267 http://arxiv.org/abs/2105.06267

[8] Olivia J Erdélyi and Judy Goldsmith. 2018. Regulating artificial intelligence: Proposal for a global solution. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society.* 95–101.

[9] Government of Malta. 2018. The Innovative Technology Arrangements and Services Act (Chapter 592 of the Laws of Malta). https://legislation.mt/eli/cap/592/eng/pdf.

[10] Government of Malta. 2018. The Malta Digital Innovation Act (Chapter 591 of the Laws of Malta). https://legislation.mt/eli/cap/591/eng/pdf.

[11] Gonenc Gurkaynak, Ilay Yilmaz, and Gunes Haksever. 2016. Stifling artificial intelligence: Human perils. *Computer Law & Security Review* 32, 5 (2016), 749–758.

[12] IAASB. 2013. ISAE 3000 (revised), assurance engagements other than audits or reviews of historical financial information. (2013).

[13] Keith Kirkpatrick. 2016. Battling algorithmic bias: How do we ensure algorithms treat us fairly?

[14] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* 33–44.

[15] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law* 7, 2 (2017), 76–99.

[16] Roman V Yampolskiy. 2013. Artificial intelligence safety engineering: Why machine ethics is a wrong approach. In *Philosophy and theory of artificial intelligence.* Springer, 389–396.