AUGUST 2022

LOOKING AHEAD

The role of standards in the future of AI governance

AUTHORS

Agustina Callegari
Ahmad Firdaus
Amaya Hana
Ana Alania
Sierra Shell
Yuri Ohkura

bsi. UCL

Looking Ahead: The Role of Standards in the Future of Artificial Intelligence (AI) Governance

CONTENTS

Please cite this report as:
Ana Alania, Ahmad Firdaus, Agustina Callegari, Amaya Hana, Sierra Shell, and Yuri Ohkura. "Looking Ahead: The Role of Standards in the Future of Artificial Intelligence (AI) Governance." University College London, 2022.

ACKNOWLEDGEMENTS

This work would not have been possible without significant support from a community of brilliant people at UCL (University College London) and beyond.

We would like to especially thank our supervisor, Dr. Irina Brass from the Science, Technology, Engineering and Public Policy (STEaPP) department for providing critical guidance, support, and encouragement through months of research. Her expertise and enthusiasm on emerging technologies and standards were instrumental in helping us navigate the difficult nature of this future-oriented research.

We would also like to extend sincere gratitude to our partner organisation BSI, especially, Matthew Chiles, Emilie Bratt, Tim McGarr, Joy Tan, Sahar Danesh, and Gill Jackson. Thank you for your trust in our capability to deliver this research project and for your active involvement throughout the process.

Finally, we would like to thank all the participants who took part in our research. Your insights were crucial for this research.

EXECUTIVE SUMMARY

The fast-evolving performance capability of AI, coupled with its broad scope of application, has the potential to bring widespread benefits to society in many domains like healthcare, education, and finance [1]. Despite these opportunities, AI is also giving rise to a number of unintended consequences, potentially leading to human rights violations, security vulnerabilities, and loss of human control [2]. Managing the scale and speed of AI adoption, and associated risks and harms, is an increasingly urgent task for governments around the world. While national AI strategies and policies have begun to emerge, regulators struggle to address the multifaceted governance challenges brought about by AI [3]. Given that significant gaps remain between regulatory protections and evolving AI risks, additional governance mechanisms are urgently needed to advance the responsible AI agenda. Standards are particularly well-placed to address the complexities surrounding dynamic technologies due to their multistakeholder nature and capacity to be reviewed and adapted more rapidly than traditional regulation [3]. This report examines the most pressing governance challenges faced by key stakeholders in the AI value chain and investigates how standards can help address their needs in the years to come.

Methodology

This research employed a mixed-method qualitative approach designed to gather forward-looking evidence on AI standardisation needs. We selected methods for their suitability to uncover the needs and challenges of key stakeholders in the AI ecosystem. Additionally, given the pace of technological and regulatory developments, our aim was to draw evidence on potential drivers of change that could reshape the AI standardisation landscape over the next ten years. To form a robust evidence base, both primary and secondary data was gathered using (1) a literature review, (2) semi-structured interviews, (3) an expert workshop and (4) horizon scanning activities.

Key findings

This report identifies several notable challenges and needs impacting the AI governance landscape.

- A lack of common language and consensus around AI risks, unclear accountability structures for risk assessment and absence of standardised measurement techniques are significantly hindering effective AI risk management efforts.
- A fragmented global regulatory landscape and overly prescriptive legislative frameworks place a heavy burden on organisations in the AI value chain, which could significantly disrupt AI trade globally. This is further exacerbated by the absence of procedural guidance provided by governments and the limited capacity of organisations to operationalise these governance frameworks.
- A lack of cross-functional skillsets, ambiguous regulatory requirements, the absence of incentive schemes and inadequate organisational structures represent the biggest barriers to responsible AI innovation and wide adoption.

While we stress the importance of a broad set of mechanisms to address these challenges, our research has found that standardisation will play a significant role in the AI governance ecosystem.

- Standards can improve trust and confidence in AI by establishing adaptive risk management frameworks, developing metrics and measurement techniques to evaluate AI system performance, and supporting conformity assessment processes and third-party audits.

- Standards can contribute to global alignment by unifying terminologies and best practices, harmonising horizontal and sectoral requirements, and supporting trade by providing procedural guidance on differing regulatory obligations.
- Process standards can enable organisations to implement responsible AI practices by clarifying accountability structures and translating ethical principles into actionable steps. By doing so, standards can promote a more level playing field across organisations, especially for small and medium enterprises.

The fast-evolving and dynamic nature of AI necessitates a forward-looking approach to anticipate and prepare for a future that is inevitably uncertain. In this context, several themes emerged which, if addressed, might promote a more future-proof standardisation process:

- Standards bodies should enhance participation from civil society, AI research, and small and medium enterprises; expand and enhance foresight methods to develop future-focused roadmaps; and reduce barriers to standards adoption like prohibitive costs and significant technical expertise required for implementation.

Finally, this report mapped short- and long-term drivers of change which could reshape the AI standardisation landscape.

- Triggers and trends in the regulatory category (e.g., kitemarks for AI) and technological category (e.g., quantum computing) were the most common drivers anticipated.

In response to our findings, this report provides the following recommendations to standards bodies:

Recommendations

1. **Contribute to the development and advancement of AI measurement, evaluation, and audit standards**
2. **Continue developing agile and flexible approaches to standardising AI and other emerging technologies**
3. **Formalise identifying AI trends and develop roadmaps for the future**
4. **Facilitate broader stakeholder representation and inclusiveness in standardisation committees addressing emerging technologies**
5. **Make AI standards more accessible for SMEs**

This report reflects a 7-month collaboration between students from University College London's Department of Science, Technology, Engineering, and Public Policy (UCL STEaPP) and the British Standards Institution (BSI). Six candidates for the Masters of Public Administration in Digital Technologies and Policy programme form the UCL student team, supervised by Dr. Irina Brass.

GLOSSARY

AI	Artificial Intelligence
BSI	The British Standards Institution
BSI ART/1	BSI Technical Committee ART 1, Artificial Intelligence
CEN	European Committee for Standardisation
CENELEC	European Committee for Electrotechnical Standardisation
EU	The European Union
EU AI Act	EU Artificial Intelligence Act (Proposal)
ETSI	European Telecommunications Standards Institute (ETSI)
GDPR	The General Data Protection Regulation (EU)
ISO	International Organisation for Standardisation
ISO/IEC JTC 1/SC 42	ISO/IEC Joint Technical Committee 1, Information Technology, Subcommittee 42, Artificial Intelligence
IEEE	Institute of Electrical and Electronics Engineers
IEC	International Electrotechnical Commission
NIST	National Institute of Standards and Technology (US)
OECD	Organisation for Economic Co-Operation and Development
SDO	Standards Development Organisation
SME	Small and Medium-sized Enterprises
UK GDPR	United Kingdom General Data Protection Regulation
US/USA	United States of America

1. INTRODUCTION

The recent expansion of Artificial Intelligence (AI) applications has benefited individuals and industry alike. Doctors can make use of decision and classification systems to improve the rate of diagnosis of diseases like skin cancer [4]. Customers seeking assistance can now reach out to chatbots or automated assistants outside of traditional working hours. A Gartner report forecasts that AI software revenue worldwide will amount to $62.5 billion in 2022, a 21.3% increase from the previous year [5]. However, the horizontal nature of the technology and the speed with which it is progressing mean that products and services are emerging before their safety, security, reliability, or trustworthiness can be assured.

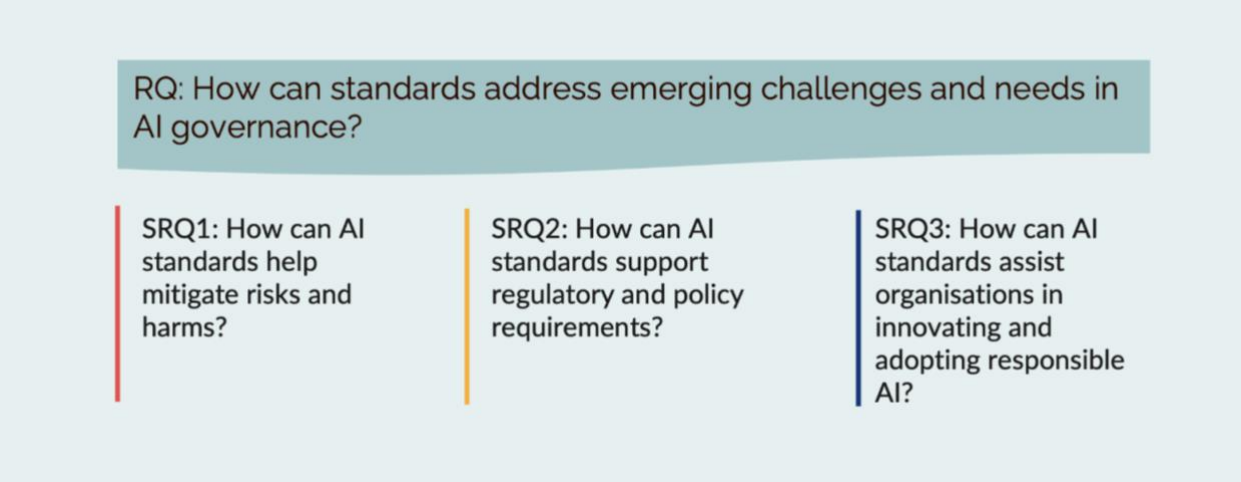
Governance is required to protect individuals, organisations, and society at large. Many tools should be considered in shaping a robust governance strategy. First, market forces are beginning to encourage responsible AI strategies; for example, 19% of large companies surveyed by McKinsey were taking steps to mitigate risks associated with explainability as of 2021 [6]. However, not all organisations are prioritising AI risk management in the absence of external scrutiny or enforcement mechanisms. As a result, governments seek to encourage or mandate responsible AI without stifling innovation [7]. Although national strategies and policies increase and mature year by year, significant gaps still exist between regulatory protections and evolving AI risks [8]. Given the lack of clarity in existing frameworks - and the complete absence of regulation in many parts of the world - additional governance tools like standards are urgently needed to support trustworthy AI development and use.

Due to their multistakeholder and adaptive nature, standards development organisations (SDOs) like the British Standards Institution (BSI) are well placed to address some of the complex governance challenges brought about by AI. However, to ensure wide use, standards must suit the emerging needs of organisations developing and deploying AI systems (bottom-up) and they must support and enable conformity with existing and forthcoming legal frameworks (top-down) [9]. Therefore, this research combines bottom-up and top-down approaches to identify the most pertinent challenges faced by AI stakeholders within industry and anticipate needs emerging from upcoming legal and policy frameworks.

Management of risks and harms associated with AI development and use will be fundamental to both approaches, as it supports the governance objectives of both industry and governments worldwide. Consequently, three research clusters have been identified to enable the collection of standardisation needs - (1) AI risks and harms, (2) AI regulation, policy and strategy and (3) AI innovation and adoption.

Furthermore, given the fast-evolving and dynamic nature of AI, it is essential that standardisation work is informed by forward-looking evidence, enabling SDOs to anticipate and prepare for the future that is inevitably uncertain. Therefore, examination of future AI governance needs is fundamental to this research, necessitating foresight methodologies to be incorporated into the design.

To guide this inquiry, the following research questions (RQ) and sub-research questions (SRQ) have been crafted to explore the role of standards in mitigating AI risks and harms, supporting regulatory and policy requirements and fostering innovation.



Even though standards cannot meet every need, our report showcases all challenges and resulting needs uncovered through a literature review, semi-structured interviews, a workshop, and horizon scanning. Due to the breadth of the project brief, and considering limited timeframe and resources, this report will not include the following: an exploration of solutions outside of the scope of standards; a full analysis of gaps within published and draft AI standards; speculation concerning the contents of standards which will emerge in two to three years; nor the geopolitical angle on AI governance.

The remainder of this report is structured as follows. Section 2 outlines the methodology used in our research process. Section 3.1 sets the scene to position standards within the AI governance ecosystem. Section 3.2 details the challenges attributed to AI risks and harms. Section 3.3 discusses

needs arising from government initiatives. Section 3.4 describes challenges faced by organisations looking to innovate or adopt AI technologies. Section 3.5 details needs which must be met to future-proof AI standards. Section 3.6 outlines drivers of change that are likely to impact the AI standardisation landscape in the future. Section 4 includes concluding remarks and analyses. Finally, section 5 provides recommendations for standards bodies, including BSI.

2. METHODOLOGY

This research employed a mixed-method qualitative approach designed to gather forward-looking evidence on AI standardisation needs. (Additional information on the methods used can be found in ***Annex A*** and ***Annex B*).** Methods were selected for their suitability to uncover the needs and challenges of key stakeholders in the AI ecosystem. Additionally, given the pace of technological and regulatory developments, our aim was to draw evidence on potential drivers of change that could reshape the AI standardisation landscape over the next ten years. To form a robust evidence base, both primary and secondary data was gathered using (1) a literature review, (2) semi-structured interviews, (3) an expert workshop and (4) horizon scanning activities (see ***Figure 1***).

Figure 1: *Research methodology structure*

2.1. Literature review

The purpose of the literature review was to gather comprehensive understanding of key debates, challenges, and conceptual issues within our three research clusters and understand the role of standards in addressing emerging industry and regulatory needs. To identify standardisation trends, the literature review was also used to map out a timeline of horizontal AI standards created by international standards setting bodies like ISO and IEEE.

The literature review was conducted using a systematic-style approach, in accordance with well-established processes and protocols commonly used in Rapid Evidence Assessments. To ensure robustness, transparency and reproducibility, we documented search strategy, syntax, database choices and inclusion and exclusion criteria. Both academic and grey literature sources were

included, however given the novelty and dynamic nature of the topic, a particular focus was given to policy documents, consultancy reports, standards and legal texts. This enabled us to gather state-of-the-art evidence from key stakeholders in the industry and keep pace with the ever-changing regulatory and standardisation landscape.

2.2. Semi-structured interviews

We conducted 21 semi-structured interviews with experts from the public sector, industry, civil society, and academia. Our intention was to (1) validate findings from the literature review, (2) gather more in-depth insights about the experiences of key stakeholders directly involved in AI development or governance and (3) draw on our participants' expertise to identify potential drivers of change impacting the future of AI standardisation landscape.

The format of semi-structured interviews enabled us to ask standard questions to all participants while providing the flexibility to follow up on interesting insights that emerged throughout the interview. Interviews were analysed using thematic analysis and an inductive approach was used to identify and examine key patterns and themes within data. To ensure anonymity, participants are referred to as 'P' and an identifying number (e.g., P12) within this report.

2.3. Expert workshop

In addition to semi-structured interviews, we organised a workshop with AI standards-makers on the future of AI standards. The online workshop was designed to enable collaborative exploration of significant events, issues and trends impacting the AI standardisation landscape over the next ten years.

In the first half of the workshop, experts were asked to brainstorm drivers of change shaping the future of AI standardisation and map them according to their impact and anticipated timeline. The second half focused on identifying and discussing the impact of these drivers on AI governance and standardisation needs and employed voting exercises to decide on priority areas for action. Future needs and drivers of change were analysed and incorporated into our findings.

2.4. Horizon scanning

To ensure the collected standardisation needs were future-facing and were informed by relevant drivers in the field of AI governance, horizon scanning methodology was employed. This helped us gather and generate drivers of change that have a significant impact on the trajectory for AI

governance in the next 10 years. Data from this activity were composed of diverse text sources, including the literature review and media articles, as well as expert elicitations through interviews and workshops. This is to ensure that the gathered and analysed drivers of change were rich in content beyond what available literature had to offer.

2. RESEARCH FINDINGS

STANDARDS & AI GOVERNANCE

3.1. The role of standards in AI governance

The global AI governance ecosystem has shifted recently from predominantly market responses to a combination of industry approaches, international frameworks, national government strategies and regulations, as well as standards development. Standards play several key roles in AI governance, accompanying other management strategies from government, industry, and beyond.

Standards support regulation. The literature confirms the function of standards in assisting government initiatives [10]. An interviewee from industry highlighted that "*standards bring enforceability and make laws implementable. They are a technical certification to implement the regulation*" [P1]. In addition, the standardisation process can be leveraged by policymakers as a tool to develop better regulation [11]. As one participant described, "*regulators are part of the standards-making process; this shows a connection with what is happening in regulation*" [P14]. In this way, standards provide a solid foundation of knowledge that can be considered and included in national and international regulations and policies [11]. Standards are also recognised in the literature as an alternative to regulation - representing an important piece in self-regulation of emerging technologies [12].

Standards codify knowledge and increase clarity. Standards systematise and classify organisational governance processes and help define common terms [13]. Due to the evolving nature of AI applications, unifying terminologies is fundamental. As one interviewee pointed out, "*standards can make high-level ideas more concrete and can help remove the roadblock and remove things that are not comprehensive*" [P1]. Through codifying knowledge, standards provide direct and detailed instructions to developers regarding how to operationalise principles [14].

Standards have a coordinating role. Due to its multistakeholder nature, standards play a coordinating role, which has been identified as crucial to collaborative innovation for emerging technology fields [15]. There is substantial agreement that cooperation is fundamental in balancing the various and sometimes conflicting interests in AI development [P11]. The standards-making

process enables the inclusion of expertise from industry, government, civil society, academia, and beyond, which would otherwise be difficult to synthesise into one cohesive document [16].

Standards help increase trust and safety. The standards-making process can help increase and maintain trust between governments, researchers, companies and consumers [14], [17]. Standards define minimum quality and safety specifications for products and services [15]. As one standards-maker asserts, "*where standards can really help…is if people feel safe to invest in them*", continuing that standards can assure customers "*this is a safe product*", which can be brought to market and generate revenue [P14]. In a field like AI which often requires technical expertise, standards can be an effective heuristic for safety and quality for customers who lack the capacity to conduct their own independent assessments.

Standards help ensure interoperability. Standards promote technical interoperability [18]. Interoperability of both data and algorithms can expand AI applications beyond their original scope. More importantly, adjusting interoperability could be a lever for discouraging the use of specific AI techniques in applications or contexts for which they were not originally intended.

3.1.1. The AI standardisation landscape

Currently, standards bodies worldwide are turning their attention toward AI. At the international level, the International Organisation for Standardisation (ISO) and the Institute of Electrical and Electronics Engineers (IEEE) are two key bodies contributing to AI standardisation efforts. National and international standards development organisations (SDOs) often work together with the goal of standards alignment. For example, ISO and the International Electrotechnical Commission (IEC) are collaborating at the ISO/IEC JTC 1/SC 42 Artificial Intelligence group to develop international standards for the entire AI ecosystem. As of August 2022, this group has published several notable standards, including those concerning the implications of broad AI governance, large-scale datasets and trustworthiness, and data quality and safety [19]. Their publication *ISO/IEC 22989:2022 Artificial Intelligence Concepts and Terminology* proved to be a milestone of 2022. Furthermore, IEEE has a large portfolio of AI standards projects, many of which address issues like general AI model representation and management, sector-specific documents like AI datasets in the medical industry, and trusted data for financial services [20].

To illustrate the work of key international SDOs, we have mapped out standardisation efforts conducted during the last four years (see ***Figure 2***). The following visualisation shows both foundational standards (those that form the basis of other standards specifications) and implementation standards (those that offer specific processes or requirements for developing or adopting AI) [21]. For example, the already mentioned ISO/IEC JTC 1/SC 42 has contributed 13 standards and is currently developing 26 additional AI-related standards [19] (for a map of draft standards, see ***Annex C***). IEEE has published three foundational and eight implementation standards.

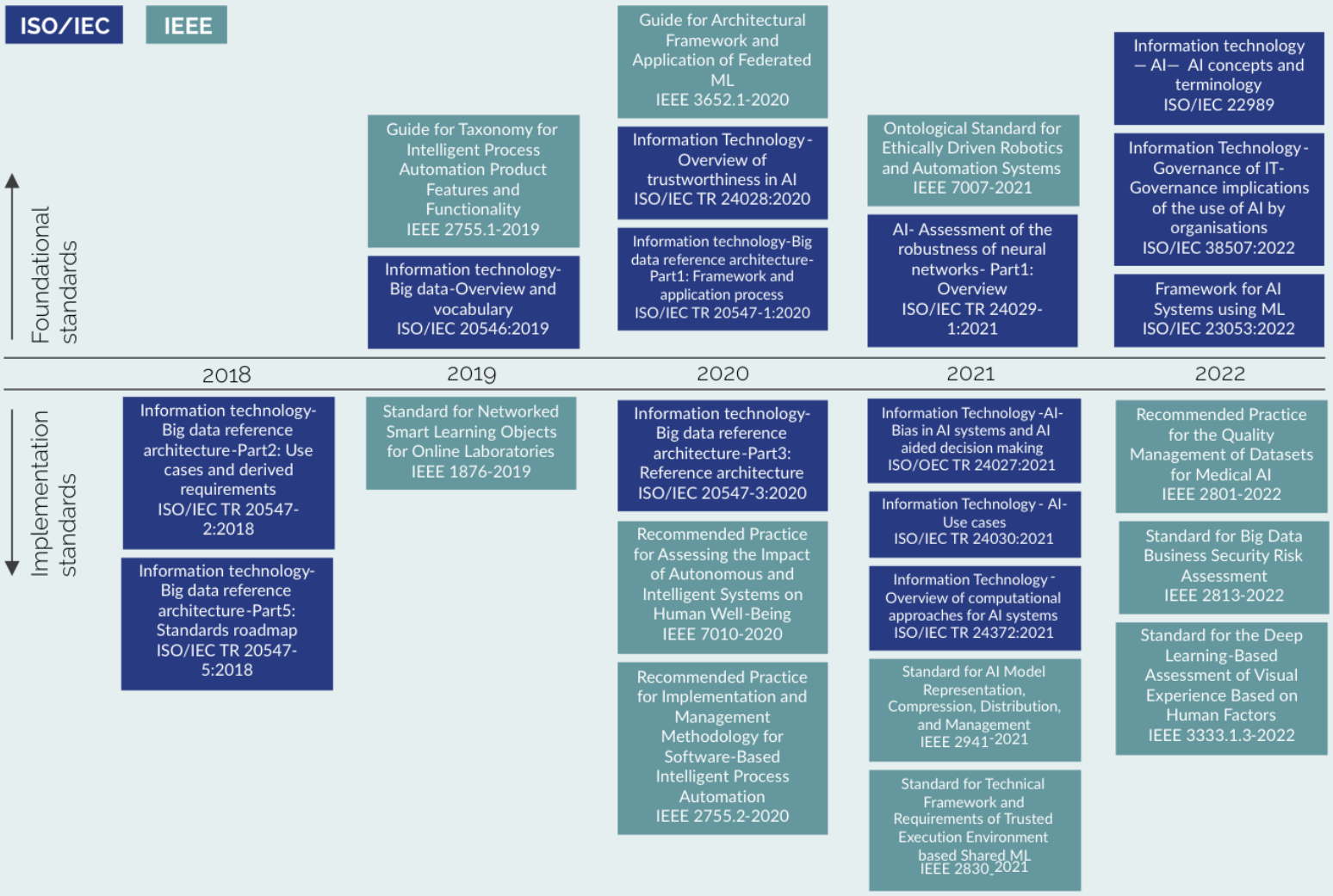


Figure 2: *AI-standards published between 2018-2022*

While not displayed in the map, it is important to highlight that, at the EU level, SDOs CEN, CENELEC and ETSI contribute to AI standardisation efforts. Other regional SDOs like NIST and ANSI in the USA and SAC in China are also active in the field. In the UK, BSI has established the ART/1 committee to discuss AI-related standards and needs, mirroring the work of ISO/IEC JTC 1/SC 42 and producing independent national standards [22]. Currently, this group has 11 published standards and 44 in progress.

Within this context, the number of AI standards already published and under development (see ***Annex C***) shows that AI is ready for standardisation [13]. AI standardisation activities are only

beginning and are likely to increase over time. However, these efforts do not come without challenges. Academics and interviewees have pointed out several critical issues that standards must address to effectively support the development and use of AI in the short and long run.

From the previous figure, it becomes clear that standards development takes time. In this regard, the literature and interviews reveal a pacing challenge which might hinder standards' effectiveness. Further, one expert argued that AI standards lack the capacity to fully address policy objectives. As one interview expressed: *"the relationship between regulation and standards has been solved in other areas, but I think it needs much more attention in AI"* [P7]

AI RISKS & HARMS

3.2. AI standards and risks and harms

Despite the opportunities of AI, this technology also gives rise to a number of serious risks, potentially leading to discrimination, privacy violations, accidents, and loss of human control [23]. AI failures, particularly in sensitive areas of application like military, medical diagnosis, the legal system, or transportation, will pose significant threats to human rights and public safety and in turn may hinder AI acceptance around the world [1], [24]. Many academics and industry experts agree that identifying, mitigating, and minimising risks will be integral to realising the full potential of AI technologies [1], [25] . As a result, a growing consensus is emerging in favour of risk-based governance of AI, centred around the necessity to assess AI risks and enable stakeholders to respond with effective and proportionate control measures.

Standardisation has the potential to help governments and companies worldwide manage AI risks and reign in its harmful impacts on individuals, organisations, and societies [9], [26]. Several interviewees emphasised the role of standards in unifying terminologies around AI risks, encouraging best and harmonised practices for risk assessment and enabling the translation of high-level concepts like "trustworthiness" into practical requirements [P4, P7, P20]. As summarised by one participant, "*standards can play an important role in unification of languages, identifying risks, measurement, and potential solutions. They can lay a path for certification programmes that are key to mitigating those risks*" [P20]. Nevertheless, several experts also stressed that AI standardisation is in its infancy and more work needs to be done to meet the challenges experienced by the industry and regulators [P7, P10, P21].

3.2.1. Key gaps and challenges

As SDOs develop their AI standardisation roadmaps, priorities should be informed by the specific needs and challenges of AI practitioners. Therefore, this section focuses on identifying challenges to managing AI risks and harms, current and future needs arising from those complexities, and the role of standards in addressing emerging needs.

Concepts and terminology in AI risk

Despite growing interest in risk-based governance mechanisms, the literature review has revealed a fragmented and incomplete body of knowledge on AI risks. Significant inconsistencies were observed from how risk itself is defined [27] to how individual risk factors like fairness, explainability and security are conceptualised and assessed [28]. These inconsistencies were also observed throughout the interviews. Although experts discussed various types of AI risks, they often addressed them in contradictory ways based on their organisational or research priorities. Reflecting on the fragmented knowledge base, one interviewee noted "*[risk] is very loaded, and people are always speaking of it in contradictory ways*" [P7], while another expert expressed concerns that "*many people in the field [of AI] are way off in their understanding of risks and harms*" [P2].

Indeed, a lack of common language and consensus around AI risks emerged as one of the central challenges faced by AI experts interviewed [P1, P3, P4, P7, P8, P18]. While recent years have seen a growing convergence on overarching principles required for an ethical and robust AI [29], several interviewees pointed to the struggles to operationalise them without a common agreement on their practical applicability and meaning. Taking explainability as an example, one expert from academia stressed that "*just showing people how we got from A to B isn't necessarily good explainability*" [P8]. They pointed to the need to "*share use cases or best practices among practitioners [to help] them understand how they can apply these basic principles when they do their business*" [P8].

In the context of financial services, one interviewee expressed that uncertainty around bias and fairness is having a cooling effect on innovation internally and is hindering collaboration with other firms externally [P18]. This is particularly concerning given AI systems are rarely built by single organisations, and often require navigating a complex ecosystem of third-party integrations [30]. An industry expert also emphasised the need to clarify 'fuzzy areas' and expressed their hopes for AI standards committees at ISO/IEC to tackle this in the next two to five years [P1].

> **Given the widespread adoption of AI and its potential for harmful impact on people and organisations, there is an urgent need to unify terminologies and clarify concepts around AI risks. This will create a foundational layer essential to enable different stakeholders in the AI ecosystem to collaborate, share best practices and agree on relevant requirements for safe, secure and trustworthy AI development and use.**

Risk assessment and accountability structures

The lack of unified terminologies and concepts on AI risks also means no universally accepted AI risk assessment model has been established to date [24]. Indeed, the literature points to the lack of a comprehensive framework and structured processes for defining, assessing and classifying AI risks [1], [27], [31]. This challenge was also emphasised by some industry experts. One interviewee

stressed that "*we are facing a problem with the risk assessment in that we don't really have a systematic way of doing it*" [P12]. Others pointed to the shortcomings of existing risk management practices to meet the unique challenges posed by AI. An expert from a large consultancy firm observed that "*organisations are looking at these risks according to their existing best practices or frameworks, which may not be fit for purpose with AI*" [P21]. This is particularly true for novel sources of risk that are unique to AI systems. One such example is the 'black box' problem, a direct result of AI model complexity [1], [27].

Furthermore, many experts highlighted the need for clarity on who is responsible for risk assessments [P7, P18] and at what stages of the AI lifecycle they should be conducted [P6, P13]. Within organisational settings, studies show that AI practitioners are not clear on who is ultimately responsible for ensuring alignment between ethical principles and product design, which significantly hinders responsible AI practices [32]. This uncertainty around accountability structures is further complicated by the fact that AI development often relies on complex supply chains of third-party data integrations, outsourcing and procurement [30].

Building trust across the supply chain has emerged as a key concern among our interviewees [P13, P18]. A researcher on medical devices stressed that open questions remain around "*liability and responsibility, especially because there are a lot of stakeholders involved in the supply chain from the developer to the medical device manufacturer*" [P13]. Other experts emphasised the importance of verifiability of claims to build trust and confidence between key stakeholders in the AI ecosystem. In the absence of such assurance mechanisms, one interviewee stressed that the risk of 'ethics washing' may be introduced: "*I notice that some people try to rank how well a company is performing on AI ethics by looking at its website. Anyone could mock up a set of principles, but it is important to see what the actual actionable activity is*" [P6].

Additionally, the literature points to the need for risk management frameworks and enforcement mechanisms to treat equally AI systems developed in-house and those procured from third-party suppliers [33]. Failing this, companies may end up outsourcing unethical projects to countries or organisations that are subject to less stringent risk assessment and compliance requirements – a practice commonly referred to in the literature as 'ethics dumping' [34].

> **Our findings reveal a need for a standardised risk management framework that accounts for risks across the entire AI lifecycle and defines clear accountability on who is responsible for risk assessments and the effectiveness of the risk control measures. To build trust throughout the supply chain, conformity assessments will be required to verify organisations' claims and certify them against a commonly accepted standard. Currently under development, ISO/IEC 42001 is an example of one such standard that will enable certification [35].**

Demand for measurement

Alongside unified terminologies and clear accountability structures, AI experts increasingly recognise the need for standardised measurement techniques to assess specific AI system risks and demonstrate compliance [36]. While many academics, industry professionals and governments have focused on defining what trustworthy AI principles should be, it is becoming evident that more targeted efforts are needed to address the 'how' question [37]. The literature highlighted the absence of measurement techniques as a notable limitation of emerging regulatory proposals, representing a crucial missing piece in the AI assurance ecosystem [37], [38]. For example, while the proposed EU AI Act necessitates transparency, it does not specify the degree of transparency needed to fulfil this obligation, or any associated benchmarks [36]. Similarly, New York City's upcoming law on automated employment decision tools requires employers to conduct bias audits but does not provide clear guidance on what metrics should be used to judge compliance [39].

Consequently, implementing organisational risk assessment processes may not prove as effective if there are no agreed upon metrics, standardised tests or measurements. This concern was shared by a researcher, who questioned how meaningful organisational risk assessment efforts could be without common requirements for quantitative measurement: "*[organisations are starting to] have processes in place to talk about risk, but they don't yet have … clear understanding of the nature of the risk they're talking about because they're going off on their own best efforts*" [P18].

In the context of standards, interviewees also challenged the usefulness of existing standards to help AI practitioners implement trustworthy AI requirements and assess performance in concrete scenarios. Speaking about robustness, explainability and auditability standards, a researcher from a technology company noted: "*These three key characteristics of trustworthy AI standards just meet the tip of the iceberg. There is nothing in them, … they never dive down into the nitty gritty or the mechanics of the topics"* [P10]. Similarly, while AI management system standards (currently under development by ISO/IEC JTC 1/SC 42) will be crucial for defining process requirements of organisations using AI, they will not define specific measurements for products to achieve these characteristics (e.g., safety or fairness) [9].

> **There is an increasing demand for tools and mechanisms that enable the assessment and measurement aspects of AI system performance like safety, transparency or bias. This will be an essential prerequisite for standardising thresholds for acceptable performance for AI risk factors in different contexts. For example, this could lead to standardising the minimum accuracy threshold required for facial recognition systems, against which they can be audited and certified.**

Dynamic nature of AI risks

Given AI's rapid development and increasing applications, AI risks are constantly changing and evolving, which further complicates AI risk assessment and mitigation efforts [1], [23]. Additionally, many AI systems are dynamic, learning during their usage and changing behaviour based on new data inputs, continued use and integrations [30]. This presents significant challenges to regulators and policymakers, as explained by an expert involved in innovation policy in Japan:

> "*Traditionally, regulators try to assess the risks of a particular technology before introducing it into practice [...] But in the case of AI, it is evolving and learning through continued use. The performance is changing, and this makes it unpredictable.*" [P8].

Concerns are emerging in the AI community that existing proposals for AI regulation are too prescriptive in their risk classifications and lack flexibility to respond to new developments in the technology [P12, P20]. One interviewee noted "*the EU hasn't really acknowledged that AI evolves, therefore some of the controls that they have put in place will quickly be obsolete*" [P20]. Several participants highlighted the need for continuous monitoring, assessments and feedback mechanisms to enable agile and adaptive governance of AI risks [P6, P8, P12, P13, P20]. These mechanisms will be especially pivotal to monitor the safety of continuously learning systems as their functionality may be significantly modified after the AI system is brought to the market. This suggests a need for almost 'real-time' conformity assessments to demonstrate compliance with relevant safety standards or legislation [9]. A prominent researcher in AI governance has further urged for frequent retesting and reassessments of robustness, bias and other measures, requiring *"a sort of flight data recorder for AI"* [P12].

> **The dynamic nature of AI necessitates a flexible and adaptive governance approach to managing AI risks. Special attention must be given to AI systems that continue self-learning after deployment to ensure that evolving risks are adequately monitored and managed throughout their lifecycle.**

AI risks are context-dependent

AI risk management is further complicated by the fact that risks arising from AI systems depend on the context of their application. As a result, the severity of consequences for not meeting responsible AI requirements varies considerably across different usage scenarios. For example, an AI system used for approving loan applications will need to prioritise fairness and explainability, whereas these characteristics may not be relevant for a recommendation system used within a music streaming app. Consequently, experts recognise that when it comes to managing AI risks, a 'one size fits all' approach is not desirable and standardised yet adaptive methodologies are needed to classify criticality of AI systems [P6, P10]. This is especially important for putting in place proportionate risk control measures and minimising barriers for innovation for non-sensitive use cases. Indeed, experts

identified determining proportionality to be one of the key challenges faced by companies working with AI systems.

> *"I think that there are risks across the board, but the most important thing is to identify what a relevant risk is [...] and then determining proportionality. Determining proportionality can be in terms of the level of assessment, the kind of verification and then possible mitigation strategies with respect to those risks"* [P6].

Risk classification mechanisms have been suggested (e.g., high, medium, low risk) to enable proportionate response measures, although experts agree that these distinctions cannot be decided by politicians, as has been the case with the proposed EU AI Act [P6, P7, P10, P12].

> **Experts increasingly indicated the need for a risk management framework that could be applicable to context specific use-cases and adaptable to different sectors. To account for contexts with different ethical sensitivities, frameworks should enable standardised classification of AI systems' criticality and identification of regulatory requirements, as well as proportionate risk mitigation measures.**

3.2.2. Standards for AI risk management needs

Although standards cannot address all AI risks independently, they will help to determine the success of AI governance strategies in protecting the public and mitigating harms.

Clarifying concepts and unifying terminologies

Standards have the potential to clarify ambiguities and build common understanding around AI risk concepts and terminologies [P3, P7, P12]. Several interviewees recognised that foundational standards, such as the recently published *ISO/IEC DIS 22989 Artificial intelligence concepts and terminology*, are important building blocks in the trustworthy AI domain as they lay the groundwork for future assurance mechanisms like conformity assessments and certification [P3, P7, P20]. Given the multistakeholder nature of AI committees, SDOs were seen to be particularly well placed to achieve consensus around key concepts such as bias or human oversight [P4].

Nevertheless, some interviewees urged caution around the role of standards in AI ethics. A government official stressed that "*quite a lot of things that people are worried about in AI risk is a genuine question of ethics or values, where people could completely disagree about the right answer*" and that SDOs are not the right institutions to set these values [P7]. Instead, standards should enable implementation of agreed-upon values proposed by governments or multilateral organisations [9].

Consequently, for AI risk areas where fundamental ethical dilemmas persist, standardisation work may face additional complexities and delays.

Enabling measurement and evaluation

To enable objective measurement and evaluation of AI system performance, several experts highlighted the importance of metrology or measurement standards [P7, P10, P18]. A standards maker from the public sector noted:

> "*I think we are making good progress on the foundational and process standards but what is next is the metrology measurement standards. How do we measure particular risks, attributes of risk or performance? I think this is what it will take for AI assurance and AI governance to be scalable*" [P7].

However, it was recognised that such standards are considerably harder to produce and will require "*focused research, collaboration and sustained testing and simulation*" over the next few years [P18]. While measurement and evaluation activities have been more common for risk factors like accuracy and robustness, more research is needed for other types of AI-related measurements like bias, interpretability and transparency [40]. Additionally, there are inherent difficulties in applying metrology to ethics. For example, fairness alone has multiple definitions and some of them are incompatible [41]. While it would be unrealistic to agree upon a single concept of fairness, "*we should at least be getting on the same page about what we mean - whether we are talking about disparate impact in outcomes or equal accuracy is how we are assessing fairness in particular contexts*" [P7].

Managing dynamic and context dependent nature of AI risks

Our findings reveal inherent challenges in AI risk management due to the dynamic and context-dependent nature of AI systems. As 'hard' legislative frameworks struggle to accommodate these characteristics, alternative governance approaches are increasingly needed to help fill the gaps.

To address the dynamic nature of AI, one expert identified the need for specific risk management standards for continuous learning systems [P12]. These standards could share best practices to ensure self-learning in AI systems is adequately moderated and the data fed back into the system goes through the right filters. Within the literature, audit trail standards have been recognised as a potential solution to improve traceability of AI systems [38]. Such standards would provide clear guidance on how to make safety-critical AI systems fully auditable and outline what documentation would be required to demonstrate compliance. We anticipate that the demand for such standards may increase over the next few years to support 'record-keeping' requirements of high-risk AI systems within the proposed EU AI Act. Indeed, a recent report from AI Watch has found that the record-keeping and technical documentation requirements of the Act are not sufficiently addressed by existing standardisation efforts, pointing to a gap [21].

Standardisation was seen as a useful mechanism by experts to account for context-specific differences in AI systems' criticality and the need for determining proportional control measures. Firstly, interviewees pointed to the need to standardise impact assessments for AI risk classification. One interviewee noted "*I think a priority topic where standardisation could bring a lot of value to different compliance regimes and businesses is a standardised impact assessment*" [P18]. Currently, criteria for such distinctions within the proposed EU AI Act is not provided, making assessments of newly developed AI systems challenging [30]. An AI expert from a prominent consultancy firm asserts that "*clear guidance will be needed in this space to make the risk-based approach functionally usable. Conceptually, I think everybody sees this as being a good approach. But it is the question of where [as an AI developer] I sit in this [classification] that will need clarifying*" [P12]. We anticipate that such standards could grow in importance if businesses are unable to determine appropriate regulatory obligations due to ambiguities around risk classifications, which would in turn impact their ability to trade within the EU [P18].

3.3. AI standards and regulation, policy and strategy

For over two decades, a 'market-first' approach to AI has been the predominant governance model. Companies, both at the national and global level, have given rise to a 'race to AI' [42], which directly impacted countries' competitiveness. However, in the last five years, increasing awareness of AI risks and harms has prompted governments to consider AI regulations, policies, and strategies. The race has now shifted into one for AI regulation [42].

Some governments are already publishing AI strategies and drafting regulations. Most serve two fundamental purposes: protecting citizens from AI risks and fostering innovation. Despite these common trends, policy and regulatory instruments vary by country. Due to the complexity of AI, countries respond differently to the question of what should be regulated and how [43]. As highlighted by the literature, AI regulation "necessitate[s] tailored policies and a holistic regulatory approach…with due attention to the interaction between the various legal domains that govern AI" [42, p. 83].

Some countries already show dominance in the 'race' to regulate AI. For example, China is one of the first countries to pass regulation concerning algorithmic governance, doing so in March 2022. In the EU, the proposed AI Act aims to position the region as the global leader in protecting society and managing risks, following their dominant approach in the GDPR [44]. The EU AI Act has the potential to become the global "gold standard" [P6], thus deserving thorough scrutiny before coming into effect.

The USA has adopted an alternative pathway, prioritising innovation above societal safeguards [45]. While the National AI Strategy is the leading document in the field, bill proposals for regulating algorithmic accountability were already presented to Congress and are expected to re-emerge [46]. The USA's structure of competing federal, state, and even municipal AI management presents a barrier to a singular, cohesive framework. So far, this has led to a decentralised and fragmented approach [47]. In the UK, the publication of the UK AI strategy and recent policy paper shed light on the country's light-touch and pro-innovation approach to AI regulation [48].

Differing political priorities have resulted in two main government approaches: risk-based and innovation-based strategies, and a combination of both. The following table describes their main characteristics and presents challenges associated with each approach (see ***Table 1***).

AI Initiative Document	EU AI Act	China's Internet Information Service Algorithmic Recommendation Management Provisions	UK's National AI strategy and its policy paper, 'Establishing a pro-innovation approach to regulating AI'	US AI National Strategy Act
Type	Regulation proposal (2022)	Regulation (2022)	Policy documents (2021 & 2022)	Policy document (2020)
Description	AI systems are subject to risk classification. There are three risk categories: prohibited AI systems, high-risk systems, and other systems [49].	Companies that provide algorithmic recommendation via the Internet shall: ● optimise their activities towards positive use of them ● not engage in activities harming national security and the social public interest [50]	The Strategy (2021) prioritises the long-term needs of the AI ecosystem, including the economy and AI governance. The policy paper (2022) proposes a set of rules based on six core principles which regulation would employ to address AI's future risks and opportunities [48].	The Act supports AI innovation with the goal of increasing prosperity, enhancing national security, and improving quality of life in the US. The Act emphasises the following policy areas: AI research, AI resources, AI workforce, and AI trustworthiness.
Main challenges & critiques	**Fundamental rights protection**: A risk-based approach could fail to protect fundamental rights [51]. **Pacing issue:** regulation takes time and the (draft) law is inflexible. The draft regulation currently provides no mechanism to label new AI systems as "high-risk".	**A narrow approach:** It applies only to "Internet information services," but does not address government use of algorithms [53]. **Privacy regulation:** it is described as a privacy regulation ,and leaves the plan for AI-specific regulation unclear [53].	**More like a signal:** The policy paper is described as a political and economic signal rather than a substantial approach for governing AI. The Strategy is described as a starting point rather than a roadmap [54]. **Brexit-related challenges:** Experts express concerns regarding harmonisation. Some participants noted that if the UK wants to	**Agency-by-agency approach:** the US's chosen approach is that specific regulatory guidelines are proposed on an agency-by agency basis [55]. This could impede predictability for AI practitioners.

AI Initiative Document	EU AI Act	China's Internet Information Service Algorithmic Recommendation Management Provisions	UK's National AI strategy and its policy paper, 'Establishing a pro-innovation approach to regulating AI'	US AI National Strategy Act
	Brussels effect: like the GDPR on data privacy, the AI Act could have a global effect in other jurisdictions and in practices of companies that want to enter the EU market [52] [P6].		retain its AI global position [P16], and the EU AIA achieves a 'gold standard' status, the UK may need to revise its current approach. **Stakeholder engagement:** Interviewees raised concerns about the lack of specificity regarding how stakeholders will be engaged in the development of AI regulation [54].	
Role of standards	Standards will support harmonisation of regulation across countries. The EU has proposed the New Legislative Framework (NLF) to align standards and regulation [26]. Harmonisation also creates a legal obligation and forces markets into compliance faster [P13].	While the law does not refer to standards, the 2017 National AI Strategy does highlight the importance of standards-setting processes in AI-driven sectors [42].	In the Strategy, the role of AI standards is to support trade; provide businesses with opportunity; ensure safety, security and trust; and support regulatory compliance.	The Initiative calls for the US standards body (NIST) to create a plan for federal engagement in developing AI standards [56].

Table 1: *AI-related policies and regulations in the EU, China, the UK and USA.*

In addition to existing regulatory and policy frameworks for AI, multilateral organisations are producing international frameworks which contribute to the current AI governance ecosystem. For example, AI principles put forth by the OECD appeared in the literature and the interviews [P3] as one of the key efforts in the field. In the international standardisation context, ISO, IEEE, and ITU standards have emerged, contributing to the AI global governance landscape.

3.3.1. Key gaps and challenges

Governing AI is undoubtedly challenging. An analysis of the regulatory and governance landscape, using primary and secondary data, has exposed several issues experienced by governing bodies and AI practitioners alike.

Lack of global alignment

Our research revealed serious concerns about the lack of global alignment disrupting AI development, access, and trade. Pointing to the fragmented regulatory landscape, one interviewee noted:

> "*There are a lot of countries that have not decided whether there should be horizontal or sectoral legislation. The EU is taking a horizontal approach, with Canada looking to head in a similar direction. Whereas the UK and USA are looking at the sectoral base*" [P12].

AI permeates many sectors, including those with established and robust regulatory environments like medicine and transport. Therefore, governments' starting points for how to best regulate AI – whether horizontally or vertically – vary. So, too, do national initiatives differ in their level of maturity; many countries have not yet developed a strategy at all.

This leaves society potentially vulnerable to known and unknown risks. A lack of global alignment, or regulatory fracture, is also detrimental to businesses. One participant points out that companies struggle and "*want to know what they need to do to satisfy compliance requirements*" [P18]. AI is an incredibly tradeable technology [P7]. However, the lack of clarity arising from global misalignment or immaturity of government initiatives creates a challenge for its tradeability. Not only are AI developers unsure of which specific requirements might be introduced in draft or underdeveloped national governance frameworks, they also lack clarity on the mechanisms and markets of compliance. One interviewee points out how a mandatory compliance mechanism like risk assessment might be even more difficult for small and medium enterprises (SMEs), asking "*how will legislation appeal to small companies or start-ups? Would they be able to run the same kinds of risk assessment?*" [P6].

> **Given the challenge global regulatory fracture brings to AI trade and international business, AI regulation needs the support of international standards bodies to spell out regulatory obligations and processes and align differing or missing national goals. Alternatively, national SDOs might model their standards on international governance frameworks. This would assist organisations in remaining compliant and being ready for new legislation, keeping AI trade global and levelling the playing field.**

Regulating a dynamic technology

Given the level of technical expertise required to both understand the current state of AI and predict its near-term development, governments may not always be the most appropriate entities to govern AI. The literature and participants reference a need for cross-sectoral knowledge so that individuals of varying levels of expertise are consulted to produce a better governance model [57], [P8, P14]. However, industry should not be solely relied upon to establish governance guidelines; literature remains suspicious of the motivation behind such private sector frameworks, as the guidelines can be self-serving, attempt to reframe a social problem as a technical one, and try to stave off regulatory efforts that may impose requirements or burdens [58].

Legislative frameworks are perceived as lacking flexibility and agility to accommodate the dynamic nature of emerging technologies like AI [P13]. Using the proposed EU AI Act as an example, numerous participants noted that amendments to its pre-defined risk categories cannot be made quickly enough, and rapid technological evolution may render these categories obsolete [P12, P13, P20]. This is also relevant because the EU's risk categories prevent adding devices which do not fit well within its taxonomy. In other words, it is not future-proof, given that not all emerging systems can be categorised within current risk tiers [30]. An expert affirmed that "*without legitimacy as to why certain systems are or are not on the red list, both public trust and the rule of law are inherently compromised*" [P13]. They also noted the need for a less restrictive framework, stating "*if you put a cap on the number of risks, once we have more risks, then you need to go to the regulation and amend… It is better to keep it open rather than keep it a closed list*" [P13].

> **To future-proof policies, regulation, and strategies, governments need access to broader technical and cross-sectoral expertise. They would also benefit from collaborating with organisations that adopt a more agile approach to AI governance, which runs contrary to governments' procedural and deliberative nature.**

"Readiness" for AI governance

There are growing concerns that countries are not ready for AI regulation. Regulatory readiness goes beyond purchasing and using new AI technology compliant with upcoming regulation. Government agencies need restructuring of crucial areas such as reconfiguring existing procedures, hiring more AI talent and refining collaborative tools to develop better knowledge sharing as new technologies emerge [59].

In the EU context, although the proposed EU AI Act has been heavily criticised, experts are confident that it will be passed [60]. However, civil society and academics express concerns that it will be immensely difficult for EU Member States to adopt the current iteration of the EU AI Act because, although the EU has legislative powers over its Member States, there is no clarity on which issues

should be dealt at domestic level and which at the EU level [61], [62]. The broad definition of AI employed exacerbates Member States' confusion [61]. The proposed EU AI Act is horizontal and faces criticism that it lacks direction for how a corresponding vertical approach might be integrated [63]. Civil society suggests various sectors should be consulted and their insights incorporated into governance models to provide clarity, as specific regulatory requirements are likely to differ across them [63].

The UK's policy paper 'Establishing a pro-innovation approach to regulating AI' rejects the EU AI Act's proposal, asserting that "we do not believe this approach is right for the UK" [48, p. 8]. The paper asserts the EU AI Act is not flexible enough, is too centralised and limits innovation [48]. The UK seeks to delegate AI regulation to existing agencies which will interpret and implement principles into AI regulation, but has not elaborated on how collaboration between agencies should be achieved [64]. Furthermore, a recent study by the Alan Turing Institute [65] reveals that UK regulators are wary that their IT systems, organisational structures and staff members are unprepared for AI which indicates a strong need for inter-organisational knowledge sharing, cooperation and skill development [65].

> **To be 'ready' for impending AI regulation, policy, and strategy, national regulatory agencies in the EU need clarity on which aspects of AI management fall within their scope of power. Globally, there is also a need for cross-sectoral cooperation across agencies for better knowledge sharing, especially concerning integration of horizontal and vertical governance. Standards can assist in regulatory readiness through harmonisation but cannot provide agencies a clear scope of power.**

Operationalising existing governance frameworks

Amidst challenges encountered with regulatory frameworks, public and private stakeholders have increasingly adopted soft laws in the form of ethical frameworks, codes of conduct and guidance documents [3], [66]. The sheer number of these documents, produced by multilateral organisations (OECD [67]), industry (Microsoft [68]) governments (European Commission's High-Level Expert Group [69]) and academia (Future of Life Institute [70]), indicates that stakeholders seek to shape 'responsible AI' conversations around their interests and priorities [37]. Normative guidelines can be reviewed and amended more frequently than traditional regulation and have been suggested to be better suited to keep pace with developments in AI [3]. However, there is growing evidence [32], [71], [72], that the impact of these guidelines is significantly restricted by their highly abstract nature and limited practical applicability. These concerns were also shared by our research participants [P3, P6, P8, P11]. One interviewee noted: "*These principles and frameworks are only as useful as they are practical. We found that people in the industry were a bit disappointed by the lack of usefulness in a lot of those frameworks*" [P6]. Both industry and government representatives agreed that the lack of empirically grounded examples and best practices are hindering operationalisation efforts. "*Currently*

we do not have enough examples to show how these frameworks work," observed a government official [P11]. Importantly, studies show this led to low adoption of ethical frameworks by SMEs and start-ups, as companies with limited resources tend to take a wait-and-see approach until use cases are more widely available [72].

Another key barrier described in the literature is the lack of adequate systems to assess and check compliance [32]. Interviewees shared this concern and stressed that without external scrutiny and assurance mechanisms, it will be difficult to "*guarantee that everything is followed or applied*" and the critical challenge over the next few years will be "*how to monitor implementation and ensure that the guidelines are complied with*" [P8]. While accountability emerged as a theme throughout interviews, experts stressed that organisations are often "*too reliant on individuals all along the AI lifecycle to enact [ethical principles]*" [P3]. Indeed, empirical evidence suggests that in the absence of formalised procedures and policies [66], as well as clear communication channels for reporting ethical issues [73], it is often individual employees that bear most of the responsibility for ethical AI. However, employees often lack adequate training or knowledge to effectively manage this process.

> **Many of the existing governance frameworks, despite their benefits, are difficult to operationalise because of their abstract nature. Empirically grounded examples and best practices are urgently needed to provide procedural guidance to organisations intending to put these principles into practice. Furthermore, assurance tools are required to enable verification of claims and demonstration of compliance.**

3.3.2. Standards for regulation, policy and strategy needs

Although not all regulatory challenges can be addressed by standards, there are several aspects of the issues identified above for which standards are well-suited.

Providing clarity in a fractured regulatory landscape

One expert describes how standards can play a key role in harmonising requirements for responsible AI, saying "*countries have different institutional set-ups, different capacities…they might regulate in different ways but following a technical standard that is stricter than all of them will help address barriers to trade*" [P7]. International standards in particular can help to address the lack of clarity and predictability introduced by a fractured regulatory landscape, as they must consider differences in national priorities. In the absence of global alignment, following international frameworks could prevent AI developers from having to retroactively redesign their products to maintain compliance with individual regions, thus facilitating trade. As one participant notes, even with regulation, "*we need actual standards… that is the way of levelling the playing field*" [P9]. Preserving fair global competition by facilitating trade is one participant's strategy for avoiding the social problems created

by Silicon Valley – "*we should orchestrate AI… so it's not in the hands of the few the same way the West Coast [of the USA] is owning our digital future. If we do that with deep tech, we have a societal problem, not just a technology problem*" [P9].

Though more narrow in scope, national standards bodies can still contribute to removing barriers to trade in a fractured regulatory space by basing domestic standards on international ones [74], with intentional deviations that support country-specific goals. Furthermore, just as global alignment will facilitate international trade of AI goods and services specifically, so might AI technologies facilitate trade more broadly. One report argues that proliferation of AI (facilitated through global alignment) will increase productivity growth which could in turn "increase economic growth and provide new opportunities for international trade" [74].

Establishing an agile approach to AI regulation

An agile standards approach could help future-proof AI governance frameworks and allow for continuous assessment. A standards expert proposed a need for greater flexibility and adaptability of the standards-making process to enable modification of standards necessitated by emerging regulatory challenges [P14]. Moreover, continuous impact assessments suit an agile approach because amendments can be made more easily. A government expert has stated that "*under the agile approach, society and stakeholders should be more generous to the idea of trial and error*" [P11]. As AI development is inherently iterative, continuous assessments are more suitable. However, another standards-maker asserts that agile standards reduce stability, and continuous amendments can be costly and time-consuming [P16]. Therefore, a balance must be sought.

Integrating horizontal and vertical frameworks

Standards can generally be classified as either horizontal and vertical. Horizontal standards apply to multiple industries, while vertical standards tend to apply to a specific sector. A policy expert has emphasised the pressing need to align vertical and horizontal standards over the next few years, asking "*how do AI standards connect with use case-specific or context-specific standards that need a lot of work?*" [P7].

SDOs organise committees which incorporate sector-specific knowledge, among other areas of expertise. This facilitates the integration of horizontal and vertical factors in two ways. First, standards-makers can create horizontal standards which take sector-specific complexities into account. Second, they can create vertical standards which complement published horizontal standards. They can do so because of the breadth and depth of stakeholders involved in the process. Though other governance mechanisms like regulation also incorporate a multistakeholder lens, SDOs can produce guiding documents more quickly than governments [P9, P11].

Furthermore, when creating vertical standards, SDOs can prioritise adaptability and flexibility within already developed vertical governance mechanisms to enable easier integration with new regulatory requirements and existing horizontal standards. If vertical and horizontal standards are properly aligned, in some cases, compliance will become easier. For example, ideal alignment of the two would mean if a technology is compliant with a vertical standard, then compliance with a related horizontal standard would be seen as an added layer of support rather than an additional burdensome obligation [P14].

Bridging the gap between principles and practice to build trust

Our findings revealed a clear and urgent need to bridge the gap between responsible AI principles and their concrete application. AI experts suggest the development of assurance mechanisms to assess, monitor and verify organisations' claims for responsible AI development and use. While a broader governance toolkit will be required to effectively translate those principles into practice, our study participants recognised that standards will play a pivotal role in sharing emerging best practice and building trust and confidence in AI systems [P6, P7, P12]. Where independent assessments are needed, internationally agreed standards will be crucial to define technical requirements needed to demonstrate compliance [P7]. One participant emphasised that audit trail standards could be used to define requirements for documenting AI development processes, saying "*we need clear audit standards around what needs to be documented, how it needs to be documented, so that they will amount to meeting certain compliance requirements*" [P12].

Additionally, experts urged that process standards, currently under development within SDOs, will be crucial for providing "*companies with the opportunity to think more deeply about what they are doing.*" Such standards can help them implement organisational structures and accountability mechanisms to consider ethical practices throughout the entire AI lifecycle [P4].

AI & INNOVATION

3.4. AI standards for innovation and adoption

Over the last decade, novel AI tools have inspired countless new products and services. AI has already altered the way many organisations conduct business. AI solutions are found in sectors like marketing, finance, healthcare, and even creative industries like music and film [75]. In the marketing sector, some literature suggests that current AI solutions boost productivity by 40% [75]. Across sectors, two of the main functions or drivers of AI adoption are to make processes more efficient and to improve decision-making [76].

Despite the potential organisational benefits of AI innovation and adoption, its expansion is not occurring evenly. Globally, of 14 regions surveyed by IBM, companies within each region are seeing an average AI deployment penetration of over 30%, with three regions surpassing 40% deployment [77]. However, in 2022, the gap in AI adoption between small and large enterprises grew; the latter are now 100% more likely to have adopted AI than the former, compared to 69% more likely in 2021 [77].

Improving adoption rates for SMEs could provide benefits beyond those gained by the organisations themselves. Limited evidence exists to assert a specific relationship between AI adoption and innovation. However, novel machine learning techniques - often developed by small start-ups [78] - could improve adoption rates by solving frequently occurring problems. Theoretically, increased adoption opens the door for innovation through sheer statistical probability.

Generating harmonised standards which solidify a common taxonomy and ensure interoperability might encourage a more level playing field for these smaller organisations developing, deploying, and using AI technologies. Exploring how standards fit within other broader mechanisms of AI innovation might help inform how to increase AI innovation and adoption around the world in a responsible manner. Furthermore, by understanding the key challenges organisations face, those challenges can be translated into organisational needs for AI innovation. Our report identifies these needs and determines which can be addressed by standards.

3.4.1. Key gaps and challenges

Organisations face barriers in the pursuit of AI innovation and adoption. An analysis of interviewee responses reveals several themes which suggest challenges are unfortunately common. Barriers discussed are also present in the literature, which validates the consistency with which organisations are likely experiencing these challenges.

Skill deficit

Experts coalesce around the value of AI expertise and risk awareness among all parties who interact with the technology throughout the AI product lifecycle. An industry expert notes that *"we have a responsibility to make sure that every single employee with an organisation is aware of [risks associated with AI] we need a cultural awareness of employee training*" [P3]. Gartner expects that by 2023, "all personnel hired for AI development and training work will have to demonstrate expertise in responsible AI" [79, para. 4].

Although the field agrees that employees along the AI lifecycle should be better educated about AI risks and best practices, experts disagree on the extent to which this is possible. Two interviewees point out the inaccessibility of the goal, especially for smaller organisations. One participant mentions that "*small companies, start-ups or even individuals cannot know everything*" [P4] and another argues that a lack of AI-related skills is a key challenge faced by SMEs [P8]. Yet another interviewee confirms that even large tech firms are having difficulty hiring experts with appropriate skills [P11].

> **To address the skill deficit, all organisation types need accessible and interdisciplinary education and training on AI risk, implementation, and management, as well as increased talent coming through the educational pipeline.**

Cross-functional expertise

To effectively create, implement, and manage AI applications, enterprises need more than one set of skills or expertise. The literature suggests that inclusivity and interdisciplinarity help foster innovation [80], [81]. Several experts emphasised areas and examples where a lack of diversity or cross-functional expertise led to a barrier or product failure. In one example, a participant from civil society recalls an AI medical device failure which occurred when the product was primarily tested by men. Female customers used the product differently, and thus preferred an alternative product on the market, leading to lost revenue [P2]. In another, a participant and policy researcher refers to a survey of Japanese businesses which revealed one of the "*major difficulties for businesses considering [AI implementation] is the lack of understanding of AI within… human resources"* [P11]. An

interviewee from a biomedical SME calls for "*a bridging person, a liaison*" to connect AI skillsets to other areas of expertise within their company.

However, one participant from a large, multinational corporation described an increase in their team's consultation with the legal department, noting inclusivity is improving "*with cross-pollination between legal ethics and AI ethics in our organisation*" [P3]. Their description suggests that at least larger, better-resourced companies are actively seeking interdisciplinarity and cross-functional skills on AI management teams.

Organisations often do not have the resources or educational materials to bring in adequately diverse and cross-silo expertise to their teams. This problem persists for SMEs which face recruitment challenges more acutely.

> **To address the lack of diversity of expertise, organisations need truly inclusive organisational structures comprised of cross-departmental skillsets, as well as education for human resources departments and hiring managers. Organisations also need access to a broad network of experts within the field and outside of their companies.**

Other internal structural barriers

The literature suggests that employees along the AI product lifecycle experience a lack of drive to change behaviour in a phenomenon known as 'behavioural lock-in' [82]. Furthermore, existing IT infrastructure is often not compatible with new technology like AI, but users still frequently choose incumbent technology [83]. Further evidence suggests that legacy infrastructure is a barrier to AI innovation [84]. One participant from academia points out that responsible AI implementation and management in particular brings additional layers of complexity that even seasoned technology firms are not used to [P19].

Given that effective organisational AI governance requires alignment of skills, resources, IT infrastructure, and beyond, a dedicated AI Operations (AIOps) team can improve operational efficiency and management [79]. A participant from industry adds that "*barriers to ethical development may be really difficult to change*", but that "*[organisations] need new communication paths for this to happen*" [P3]. Communication pathways can be formalised through AIOps, as well as a transition away from legacy infrastructure and social reluctance. However, some smaller organisations are unable to prioritise a dedicated AIOps team. As an interviewee from a large multi-national company puts it, "*huge companies pay their employees better and therefore have more power to hire talented people*" to fill operational positions [P11].

> **To address internal structural barriers, organisations need dedicated teams to strategise, synthesise, and overcome those barriers. Companies could use management systems to streamline technological and social transitions, widespread adoption of AIOps, and resource pools (financial and technical) for smaller organisations operating with a leaner structure.**

Lack of clarity and predictability

Several interviewees explicitly mentioned how a lack of predictability and clarity in the AI governance landscape creates barriers for organisations [P10, P11, P12, P18]. As a participant from the private sector states, *"when it comes to principles and practices, organisations are quite blurry"* [P10]. According to interviewees' comments, the lack of clarity derives from ambiguity in regulatory language, requirements for bias testing or mitigation, and even difficulty in identifying the appropriate standards to adopt. A standards-maker asserts that "*uncertainty has a chilling effect on innovation internally, and on collaboration with other firms*" due to a lack of an accountability structure [P18]. They continue, emphasising a need for balance between clarity and rigidity. They note sometimes businesses prefer clarity for satisfying compliance requirements, but sometimes they argue frameworks are too prescriptive – "*sometimes businesses argue for [the former] to remain compliant, but other times they say the opposite because they'd rather just do what they're currently doing, as that's the easiest*" [P18]. Some organisations are unsure how to operate, given a lack of clarity that plagues the regulatory landscape of emerging technology. Others use the information vacuum as an excuse to advocate for more or less prescriptive governance.

> **When presented with government initiatives that lack specificity, thus impacting predictability, organisations need direction in the form of strategic pathways which include management frameworks and process standards. Literature supports a role for expanding the private regulatory space [85] in helping organisations navigate strategic pathways for responsible AI and validate those pathways through assurance mechanisms.**

Prioritisation

Radical transformation such as AI adoption entails set-up and other fixed costs [86], and it is crucial for companies to have sufficient financial resources and human capital. The lack of financial resources can result in deprioritising the transformation of outdated system infrastructure and novel technology deployment. This is also called 'institutional lock-in' [86], and SMEs tend to experience its impacts more frequently due to their insufficient capacity [83], [84]. A survey [84] conducted in

the UK reinforces the significant difference in AI adoption between SMEs and large companies, and institutional lock-in might be a culprit. Beyond the literature, three participants [P3, P6, P18] specifically brought up prioritisation of resources as a key barrier organisations face in delivering responsible AI. In summary, they assert that companies with fewer resources must prioritise profit-generating features requested by customers over features of responsible AI, or else take cost-saving institutional measures, like keeping headcount low.

> **Responsible AI innovation is often deprioritised due to financial or customer pressure to prioritise revenue-generating features, or due to difficulties in changing legacy IT infrastructure and human behaviour. Given such inevitable restrictions, regulatory pressure is needed for a universal focus on responsible innovation. In conjunction, financial and social resources could assist SMEs in meeting regulatory requirements, given their more pressing need to prioritise growth.**

Incentives

Industry is inherently revenue-driven. As one standards maker puts it, R&D investments - including in responsible AI - are "*expected to generate revenue in the future*" [P10]. In some cases, this creates a misalignment of incentives when profitable products or customer partnerships reject or deprioritise more ethical AI tools. For example, an expert from industry commented that often, revenue and "*doing the right thing*" are in conflict [P3]. They suggest industry "*might require a different incentive model to promote ethical behaviour and product innovation*" [P3].

Still, not all experts or literature agree on this point. In some ways, incentives are indeed aligned with improving responsible AI practices. An expert from civil society points out that "*over 80% of AI projects fail due to moral or ethical issues or liabilities"* [P2]. Another expert commented that "*the cost of failure in this area is high*" [P6]. Solving these issues can help get products to market more quickly and reduce failed pathways, a goal aligned with industry's prioritisation of revenue. Multiple experts [P7, P3] note that trustworthiness or data ethics features are being requested of their products by their customers, pointing to some market demand for responsible AI. The former interviewee argues that a driver for adoption of assurance processes comes from "*consumer demand and competition on the basis of trust*" [P7]. The literature supports this assertion as well, calling responsible AI a driver of near-term innovation [79].

However, one participant comments that "*most companies aren't going to have an incentive to take this seriously until there is legislative intervention*" [P6]. The evidence suggests that, even where inclusion of certain responsible AI features is incentivised in the market, not all shortcomings of AI can be addressed in this way, nor by all organisations. To offer a solution, one participant from industry describes the need for a "*fortitude of leadership… having the conviction to say… 'even if customers are willing to pay for these systems, we will not use the technology for this purpose'*" [P3]. Not

all firms will be able or willing to make such choices, especially SMEs who often cannot afford to be so discerning.

> **In the absence of consistent and adequately robust incentives for responsible AI innovation, organisations need new incentive models, perhaps similar to the use of "bug bounties" in the cybersecurity field, which would provide financial rewards for identifying AI risks. Organisations also need government mandates and increased awareness by customers and consumers of the benefits of certified or otherwise responsible AI products and services, driving demand.**

3.4.2. Standards for industry needs

Not all challenges organisations face are best addressed through standards. They should not be seen as a panacea of AI governance for organisations, but instead should work in conjunction with other mechanisms like government initiatives, hiring practices, and internal training. Regardless, standards can address many of the challenges identified by our research to some extent.

Addressing internal barriers with process standards

One secondary goal of AI standards - beyond securing product safety, trust, and regulatory support - is to help firms which create complex technology function more effectively. As one standards maker argues, "*in the long run, [standards] should help [companies] work more productively…because you're putting the right…procedures in place…then you can avoid AI model failure*" [P18]. Management systems will play a key role in helping organisations restructure their operations to ensure responsible AI and find clarity in an evolving governance landscape. One participant notes that process standards "*which incorporate elements of risk management or quality management should be a focus of standards bodies moving forward*" [P7]. If process standards are thorough enough, they can help meet the need of reducing internal structural barriers when managing and adopting AI. Many standards bodies are already pushing management systems, such as ISO's forthcoming 42001 which is "intended to aid organisations to develop or use AI responsibly" [24, p. 24].

However, process standards, and standards in general, will only be as effective as they are accessible by all organisation types. Two experts assert that a large majority of AI businesses in their respective home countries are SMEs [P9, P18]. Therefore, many of the potential adoptees of standards are functioning with limited resources, smaller teams, and amidst difficult prioritisation decisions. Standards have the potential to level the playing field, or at least improve the landscape for equal contribution to AI innovation. An interviewee argues "*there is an inherent barrier [for SMEs] that cannot be passed without standards intervention*" [P9]. To do so more effectively, standards bodies could actively recruit representatives from SMEs into standards committees.

Expanding the access and expertise of SMEs

The literature suggests that some standards-making processes are not sufficiently inclusive, with one report stating "both SMEs and consumer associations argue there should be more steps taken… to enable SMEs and consumers of all ages and capacities to participate more effectively than at present" [26]. Better integration between standards-makers and SMEs could benefit both groups.

SMEs would gain access to mentorship, networking opportunities, and cross-sectoral expertise found on standards-making committees. As one paper asserts, "within companies, closer connections between product teams with experience in standards and AI research teams can spread this culture [of responsibility and safety]" [17, p. 32]. This suggests that integration between the two groups could potentially address even behavioural lock-in described in the literature. One participant argues that barriers SMEs face due to impending EU legislation require "*interventions, support, and mentoring*" [P9] which could be facilitated by SME membership in standards committees. Furthermore, in addressing a lack of cross-functional skills, SMEs would have closer access to the breadth of expertise a standards committee contains, beyond technical experts and ethicists and including "*people from the music, finance, and medical sectors*" [P14].

In return, SDOs would receive input from a large segment of the market who might consequently purchase standards more frequently. In addition, better integration with SMEs could act as an investment in standardisation more broadly; the EU Parliament has published in a roadmap document, "engaging the research and innovation community early on in standards development also provides an opportunity to build expertise and skills in standardisation" [87, p. 8]. Given the unique challenges of standardising digital technology which evolves rapidly, such an investment could benefit the entire field.

Providing clarity and predictability

Standards help organisations by alleviating confusion and improving predictability as new government initiatives emerge, protecting trade and thus enterprises' market access and revenue. Our report provides a detailed overview of how standards meet this need (see ***Section 3.3***).

Barriers to standards adoption

Finally, many organisations cannot take full advantage of the benefits of standards. They face barriers like a lack of technical expertise, limited financial resources, or difficulty understanding which standards are appropriate for their use case. We consider these and other barriers to adoption organisations face (see ***Section 3.5***).

3.5. Future-proofing AI standards

Our report has revealed many current challenges related to AI risks and harms, government initiatives, and innovation that can be addressed - at least in part - by standards. Building on literature and primary data collected, we map out a set of future-focused challenges and needs that, while applicable to other emerging technologies, are of great importance when future-proofing AI standards.

3.5.1. Broadening the standards-making community

Standards-making involves many stakeholders. Indeed, other governance mechanisms also utilise a multistakeholder approach. However, as the multistakeholder approach is fundamental, it should be prioritised across all spheres of governance. Therefore, it is valuable to outline how to increase and improve relationships between stakeholders, including within SDOs. WIth this goal in mind, we identified the following needs:

SDOs should expand inclusion of civil society and the research community. Standards are not created only by technical communities. SDOs do actively and successfully seek expertise from a wide variety of sectors [P13]. Nevertheless, interviewees highlighted several areas for improvement. Three participants mentioned the importance of non-technical stakeholders in standards-making [P1, P7, P10]. Furthermore, academics recognise a mismatch between stakeholders currently included in the standards-making process and those involved in the creation and deployment of AI [26].

Three interviewees from government, academia, and industry explicitly mentioned civil society participation as one area for improvement in the multistakeholder model [P1, P4, P7]. As an expert from the private sector observed during an interview, "*there seems to be a gap in the representation of civil society*" [P1]. In addition, the literature supports the importance of engaging civil society in AI standardisation, highlighting that AI standardisation approaches must consider ethical and societal concerns, and this needs to be done by including civil society in the process [8].

Our research also indicates the need for early engagement with the research community in standards development. Currently, AI research organisations, like labs, are absent from ongoing standardisation efforts, which could negatively impact the relevance of standards [17]. The European Commission has underlined that the research and the innovation community do not regard standardisation as a priority due to a lack of resources; "they consider that time spent on standardisation activities is not sufficiently rewarded" [87, p. 8]. However, standards bodies need to increase their attention to these communities. As another participant from academia mentioned, research communities help drive the AI research agenda [P4]. Consequently, if the research community is involved in standards-making, both AI research and standards processes can benefit.

Finally, our research clearly points to a need for better integration of SMEs in SDOs. Key to novel AI innovation, small organisations like start-ups would benefit from having their perspective included in the standards-making process. Furthermore, SDOs might broaden the base of standards adopters as well, if subsequent standards are more accessible to SMEs due to their inclusion in the process.

Coordination and communication between SDOs could be improved. Several standards bodies worldwide are concurrently working on AI-related norms and processes [13]. While coordination across bodies has improved in recent years, the collaboration process itself was described by a standards-maker as "*difficult*" [P7]. The need for closer alignment between SDOs, especially between national/regional bodies and international ones, appeared clearly in the interviews with participants from academia and government [P7, P10, P14]. Better alignment might mean avoiding duplicating efforts [P7]. For instance, an interviewee from academia stressed that "*BSI already has this alignment with ISO, but I think from a multinational perspective, it is helpful to have them plug into and emphasise the alignment with standards that are doing things at ISO level*" [P21]. To improve collaboration, standards bodies must review the structure of their partnerships [P7]. Facilitating a global platform for collaboration on standards development was also suggested by an interviewee: "*we do need to have that international understanding of what we can achieve to really make standards useful and also to make sure that it has the right level of impact*" [P14].

3.5.2. Foresight methods to develop future-focused roadmaps

To future-proof AI standards, SDOs should continue evidence-based research and incorporate additional foresight methods. By doing so, standards-makers can create future-focused roadmaps, improved through regular iteration and adjustment. They can also guide prioritisation for governance of AI, a technology which is inevitably subject to disruption by other emerging technologies.

Iterative standards-making allows for continuous review and improvement. The importance of reflecting on the performance of each standard was highlighted by a participant: "*[Standards bodies] should be prepared to say what worked and did not work rather than discuss what should and should not*

be included. It is important to know what did not work to improve and identify why" [P6]. SDOs can contribute to research suggesting how to best govern emerging technology in the long term while making adjustments in the short-term, helping governments develop an evidence-based strategy. In the words of an industry expert, "*standards bodies should be motivated to develop as much of the empirical basis and evidence for that to emerge*" [P6].

The AI standardisation process should include identifying trends and roadmapping for the future. Interviews revealed the need to ensure relevant processes are in place to identify future standardisation needs. Specifically, participants mentioned observing trends and roadmapping future AI standardisation needs [P10, P12]. An industry participant pointed out that currently, "*we don't have that many evangelists in standards industries who look at trends, looking at five years, ten years, fifteen years. Standards that I see today are focused on the now. They're looking at a maximum a few years from now*" [P10].

Analysing trends is important, however such analysis is only useful if it is incorporated into SDOs' agenda. One output of this process might be creating roadmaps which capture observed trends and even drivers of change. While roadmaps laying out standards' role in supporting emerging technologies have already been implemented [16], continued use of this instrument is encouraged.

3.5.3. Balancing agility and diligence

The AI standardisation process should allow for agility. The pace of AI evolution presents a challenge to standardisation. One participant expressed concerns about how standards cope with emerging technology, noting "*the standardisation cycle takes a long time ...even one sentence can take up to a year*" [P1]. Due to the laborious nature of standards-making, those which address social and ethical issues are typically only developed once relevant problems are mainstream [16]. Indeed, one UK government report suggests that the cadence of standards-making renders it inappropriate for fast-developing technology like AI [15]. Whether or not this is accurate, improving the agility of SDOs might ease the process of standards amendment when new applications and techniques appear. Standards need to respond quickly to the number of issues that AI technologies bring. Still, standards development is already typically more agile than developing governmental strategy.

However, agility cannot come at the price of diligence. One standards-maker points out that, in many cases, the standards development process is intentionally slow and deliberate [P16]. The process, which seeks consensus, requires coordination, collaboration, and compromise between stakeholders with different interests. This inherently takes time and cannot easily be disregarded in favour of speed. Concurrently, as a standard-maker highlighted, the complexity of the process provides stability to those who need to implement these standards [P16]. Nevertheless, it was also recognised that "*there are sometimes delays that are not necessary*" [P16]. There are ways to improve

and shorten the process: "*It is clear that there is a need for agile standards, but there is also a need for stability. It is important to get that mix right*" [P16].

Flexible standards models might offer a balanced alternative. We acknowledge that our partner organisation BSI is already conscious of and taking measures to address this issue. BSI Flex Standards, described as a "fast-track standard" [15, p. 42], are designed for emerging technologies which bring a high level of uncertainty. Their iterative nature allows stakeholders to see how standards might be implemented and suggest changes in accordance with their needs. Other standards bodies could look to the Flex model when balancing agility and diligence.

3.5.4. Removing barriers to standards adoption

Barriers to standards adoption emerged from two different but interconnected angles: a lack of adoptee expertise and prohibitive cost, especially for SMEs.

SDOs should recognise and address the high level of technical expertise required to adopt certain standards. Interviewees highlighted the importance of lowering the level of expertise standards required for implementation and in navigating certification and compliance structures. A standards maker commented that they were "*surprised by the amount of technical expertise required in standards*" [P1]. One potential solution put forth by a participant was using more accessible language: "*Plain language makes a huge difference in whether or not people can take action*" [P19].

Going beyond understanding the contents of the standards themselves, organisations do not always understand what questions standards will answer for their business. This is especially true for smaller organisations; as one participant notes, "*most start-ups have problems understanding what standards are…and finding the right standards*" [P12]. Though SMEs suffer acutely from this problem, larger organisations are not immune. A standards-maker remarks that, even for companies that have legal, conformity, or quality assessment teams which could help implement standards, "*that doesn't always connect the dots [or address] everything that needs to be rolled out*" [P14].

One way to increase the accessibility of standards is to improve the interface design and overall user experience. This approach is also supported by the European Commission, which stressed the need for future standards to be more user-friendly, particularly for SMEs [87]. In addition, SDOs could engage in further research to understand how to best lower the level of expertise standards adoption requires.

SDOs could selectively reduce the costs of standardisation. When considering the financial cost of standards as a barrier, the need for balance is clear. Standards bodies must undoubtedly fund their activities in some way. It would be unreasonable to suggest that all standards should be free and open-source, or that certification services should be free of charge. Still, improving the availability

of certain standards or assurance processes was suggested by one participant from academia, who recommended making certain standards "*free and openly available*" [P21].

Nevertheless, a standards-maker explains that the SDO business model might not support completely gratis standards, especially given that more balanced approaches are available [P16]. In one example, the European Commission calls upon European standards organisations to "integrate open source solutions into their activities, which can provide SMEs with quick (...) solutions in the uptake of technological solutions" [87, p. 8]. Another example could be providing standards for AI audit trails for free, encouraging responsible innovation [17]. Standards bodies could develop and leverage their relationship with the compliance and certification industry to negotiate the right balance between supporting start-ups and maintaining revenue.

DRIVERS OF CHANGE

3.6. Drivers of change could reshape the AI standardisation landscape

Standards can address many of the challenges uncovered in the literature and primary data collected. However, our research has made clear that, in some instances, current AI standardisation needs are not being fully met. In order to reduce these gaps moving forward, as a proactive measure, SDOs should consider future needs when creating standards for rapidly evolving technologies. This is especially true considering AI could change upon interaction with other 'disruptive' emergent technologies.

To generate future needs, we consulted with experts to determine what drivers of change are likely to impact the AI standardisation landscape. Through our research, we collected 47 drivers of change within five categories. Referencing or tracking such drivers can ensure standards produced are relevant, useful and timely for AI practitioners. As literature has pointed out that standards in this area can easily be overtaken by newer technologies [88], SDOs would benefit from incorporating foresight into their process.

Drivers of change have been plotted according to a timescale of the most likely period in which the driver is likely to occur, intensify, or have the greatest impact (see ***Figure 3***). Drivers are colour-coded to show their source of origin, suggesting that a workshop which targets driver generation might be the most efficient way forward for standards bodies seeking to incorporate future-focused methods into their work.

Figure 3: Drivers of change map

The Regulatory/Governmental category shows the greatest number of drivers, which reflects the moderate maturity of AI as a technology. Though new techniques and applications are developed every year, AI is not an entirely new field, and AI systems exist in many products and services already on the market. However, governance mechanisms have not yet caught up to AI proliferation. Therefore, it makes sense that regulatory or governmental drivers will grow in the short term, as evidenced by the large number of drivers plotted in the '1-2 years' ring. This aligns with the pattern of continuously emerging regulation and national strategies.

Furthermore, the literature suggests that technological innovation (like the advent of quantum computing) can beget further innovation [89]. Technological advances must be considered in understanding how AI's capabilities may evolve, especially given its horizontal nature. As other horizontal technologies like Web 3.0 or battery technology evolve, understanding interactions between them becomes fundamental. Therefore, it is not surprising the Technology category is also heavily populated.

Most drivers of change are likely to occur or intensify in between one to three years, with a total of 34 plotted. This is not surprising, given that shorter-term drivers are likely better understood and more predictable than long-term drivers of change. Furthermore, drivers which are more top-of-mind for participants, like those currently envisioned but not yet deployed (e.g., Web 3.0 [90]) might appear at a higher rate due to the availability heuristic.

Though a full analysis of all drivers of change is outside of the scope of this report, by examining how a single driver of change - like 'publicised AI safety incidents' - can impact the landscape, we hope to showcase the value of tracking and considering them when creating standards.

Zooming in: 'Publicised AI safety incidents' as a driver of change

Publicised AI safety incidents already occur. For example, in July 2022, a chess-playing robot broke the finger of a young boy in a chess tournament when he moved his piece out of turn [91], [92]. However, as AI adoption grows in the near term, newsworthy stories are likely to emerge at a greater rate about various harms, increasing the public's awareness of specific AI risks. The public may lack the nuance of which aspects or applications of AI are safe and demand restrictions, even in instances where AI is harmless or helpful.

Given public demand, governments may adopt more stringent regulations, leading to a more pressing need for immediate designated standards. It is likely that there will be an escalation in the development of standards requiring risk or impact assessments, or standards requiring certification marks (e.g., BSI Kitemark or CE mark) for transparency.

Alternatively, perhaps public trust in AI might deteriorate to a point that companies may no longer invest in AI development or deployment at the same rate as they currently do, regardless of regulation. This would change the priorities of SDOs which might shift focus away from AI.

Collecting, examining, and tracking drivers of change can assist SDOs in being proactive about the future of the standardisation landscape. By including a detailed plot of drivers, our research hopes to provide a baseline for this process and encourage further research both on specific drivers themselves and around additional drivers not uncovered in our research.

4. CONCLUSION

AI is a powerful technology that has the potential to transform many industries and bring widespread benefits to people and organisations. However, as this research has demonstrated, appropriate governance mechanisms are urgently needed to ensure that AI systems are developed responsibly and potential risks and harms are effectively managed. As policymakers, standards bodies, and industry embark on creating and implementing such mechanisms, they inevitably face challenges due to the inherent complexities and uncertainties surrounding this technology. This research aimed to uncover the most pertinent challenges and needs within AI governance and establish how standards can help address them in the years to come.

Section 3.2 revealed that a lack of common language and consensus around AI risks, unclear accountability structures for risk assessment and absence of standardised measurement techniques are significantly hindering effective risk management efforts. Challenges with risk-based AI governance were found to be further complicated by the dynamic and context-dependent nature of AI systems. Section 3.3 uncovered that a fragmented global regulatory landscape and overly prescriptive legislative frameworks place a heavy burden on organisations in the AI value chain, which could significantly disrupt AI trade globally. This is further exacerbated by the absence of procedural guidance provided by governments and the limited capacity of organisations to operationalise these governance frameworks. Finally, Section 3.4 demonstrated that a lack of cross-functional skillsets, ambiguous regulatory requirements, the absence of incentive schemes and inadequate organisational structures represent the biggest barriers to responsible AI innovation and wide adoption.

While we stress the importance of a broad set of governance mechanisms to address the above challenges, this research has found that standardisation will play a significant role in (1) building trust and confidence in AI, (2) improving global alignment, and (3) enabling organisations to implement effective organisational processes needed for responsible AI development and use.

Standards could improve trust and confidence in AI

To increase AI adoption and encourage innovation, businesses developing and using AI need assurance mechanisms to demonstrate responsible AI behaviour to regulators, consumers, and one another. Our findings point to specific standardisation activities needed to help meet this objective over the years to come. First, an adaptive risk management framework is required to effectively manage evolving AI risks throughout the entire product lifecycle. Second, standardised and reliable measurement techniques are needed to assess specific AI system risks and evaluate performance. Continued research is necessary to ensure that

relevant metrics, metrology standards and evaluation methods are developed to assess important characteristics like transparency, security and fairness. Finally, to enable organisations to demonstrate the trustworthiness of their products, third-party audits and other conformity assessment processes are increasingly required. Audit trail standards can define requirements for documenting AI development across various applications.

Standards could contribute to global alignment of AI

In the face of a fragmented regulatory and policy landscape, standardisation has emerged as a key instrument for tackling global alignment in AI governance. As international standards are implemented across multiple jurisdictions, they have the capacity to unify terminologies and spread best practice regardless of differing or missing national regulatory requirements. Additionally, standards aligned to specific governance frameworks, like the EU's proposed AI Act, could assist international businesses in remaining compliant and continuing to trade globally. Standardised risk management frameworks and impact assessments were found to be particularly needed to help achieve this goal.

Standards could enable organisations to implement responsible AI processes

Standardisation efforts can support the creation of processes and management frameworks to address organisations' structural barriers to responsible AI innovation and adoption. Standards can also reduce the burden that AI regulations and policies sometimes generate. Finally, standards can help organisations define and abide by clear accountability structures and steps, reducing uncertainty about how ethical principles should be applied across the entire AI lifecycle. By doing so, standards can promote a more level playing field across organisations, especially for SMEs. SDOs could improve how they integrate the SME perspective into the AI standardisation process, providing informal access to AI expertise and potentially improving the relevance of AI standards for small firms.

Given the speed of AI technological development and its growing presence in the market, SDOs should identify current AI standardisation challenges and future needs in a timely and comprehensive way.

Concretely, for standards to be future-proofed, our findings suggest the need for strategic iteration and revision of AI standards, strengthening agility in specific areas. Flexible standards offer a model which balances diligence and agility. Standards bodies should use foresight methodologies to advance and improve roadmapping. For example, by collecting, analysing, and tracking drivers of

change, SDOs can identify events or trends that could impact AI standardisation processes in both the short- and long-term.

Limitations and future research

Given the broad scope and horizontal nature of this research, we were unable to provide a fully comprehensive analysis of the three research clusters selected. For example, while we explored challenges related to assessing and mitigating AI risks, it was outside of scope to delve deeper into specific risk factors like explainability and data quality and explore whether current and emerging standards are addressing them adequately. We believe that such gap analysis would be extremely valuable for organisations like BSI and should be considered for future research. Additionally, a more comprehensive standards-mapping activity could incorporate both horizontal and vertical standards relevant to AI and identify areas of alignment needed between them, as suggested by one of our research participants. Lastly, examining the geopolitical implications of AI standardisation would be valuable. While this report refers to several countries' regulatory and policy initiatives, it contains limited insight into differences in national approaches to standardisation itself.

5. RECOMMENDATIONS

1. Contribute to the development and advancement of AI measurement, evaluation, and audit standards

Given the growing demand for AI assurance tools, techniques and processes, participants in this research stressed the urgent need for measurement and evaluation standards to enhance the auditability and traceability of AI systems. Therefore, to ensure the UK can continue to take international leadership in standards development, we recommend BSI to consider exploring the following standardisation areas:

- Develop metrics and measurement methods to assess performance of AI systems, particularly for underdeveloped areas like explainability, data quality and fairness;
- Identify standards needed to support algorithmic auditing and certification; and
- Develop audit trail standards to define requirements for documenting AI development across various applications.

2. Continue developing agile and flexible approaches to standardising AI and other emerging technologies

Standards development processes must account for both the dynamic nature of AI and the lengthy time required to achieve consensus between key stakeholders. We recommend BSI continue and expand its work on agile and flexible approaches, such as Flex Standards, as follows:

- Explore a planned adaptive approach to promote new knowledge generation and incorporate such knowledge into a codified revision process [93]. One way this might be accomplished is through a formalised review of the standards review process itself, conducted with relevant experts who are not part of BSI committees. Through a similar procedure, the metrics by which BSI committees assess the effectiveness of standards themselves might be regularly reviewed by external experts.

3. Formalise identifying AI trends and develop roadmaps for the future

Many SDOs are likely already aware of future-focused research methods. Still, foresight research methods can allow BSI to be proactive when creating standards. Formalising the process of identifying potentially disruptive trends or triggers will help future-proof standards, especially those which apply to rapidly evolving technologies like AI. By formalising such methods, BSI is more likely conduct the following with regularity:

- Horizon scanning, which brings to light weak signals of change that are pertinent to AI development [94];
- Driver mapping, which plots signals of change by time frame and category [94]; and
- Scenarios, which help to explore alternative futures within a particular area, like AI [94].

4. Facilitate broader stakeholder representation and inclusiveness in standardisation committees addressing emerging technologies

The importance of multistakeholder cooperation in AI standards-making has been underlined in both the literature and interviews. However, this research has revealed moderate gaps in the representation and inclusion of specific groups. BSI can expand and facilitate diversity and inclusion in the AI standardisation process by actively recruiting to their committee representatives from:

- SMEs;
- Civil society; and
- Industry research, such as AI research labs.

5. Make AI standards more accessible for SMEs

Our research indicates two significant barriers to standards adoption that SMEs face more acutely - a steep learning curve in understanding and adopting AI standards and the sometimes prohibitive cost associated with compliance. BSI should implement a balanced solution that includes the following:

- Integrate open-source solutions into their activities where possible, which can be achieved through coordination with government initiatives to support SMEs;
- As a short-term solution, leverage their close relationship with the government to lobby for a scheme that provides a financial pool for SMEs to adopt standards; and
- Develop long-term solutions which do not make SMEs reliant upon external subsidies, but still facilitate their use of standards in key areas like AI risk management.

REFERENCES

[1] B. W. Wirtz, J. C. Weyerer, and I. Kehl, 'Governance of artificial intelligence: A risk and guideline-based integrative framework', *Gov. Inf. Q.*, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126090396&doi=10.1016%2fj.giq.2022.101685&partnerID=40&md5=ba7bab1e031c65b0fd444ce43f61599f

[2] B. Cheatham, K. Javanmardian, and H. Samandari, 'Confronting the risks of artificial intelligence', *McKinsey Q.*, vol. 2019, no. 2, 2019, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079283855&partnerID=40&md5=86169ee128e7c7f33dbf9659a97fda36

[3] A. Taeihagh, 'Governance of artificial intelligence', *Policy Soc.*, vol. 40, no. 2, pp. 137–157, Apr. 2021, doi: 10.1080/14494035.2021.1928377.

[4] R. C. Maron *et al.*, 'Artificial Intelligence and Its Effect on Dermatologists' Accuracy in Dermoscopic Melanoma Image Classification: Web-Based Survey Study.', *J. Med. Internet Res.*, vol. 22, no. 9, p. e18091, Sep. 2020, doi: 10.2196/18091.

[5] Meghan Rimol, 'Gartner Forecasts Worldwide Artificial Intelligence Software Market to Reach $62 Billion in 2022', Gartner, Press Release, Nov. 2021. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2021-11-22-gartner-forecasts-worldwide-artificial-intelligence-software-market-to-reach-62-billion-in-2022

[6] Daniel Zhang *et al.*, 'The AI Index 2021 Annual Report', Stanford University, Mar. 2021. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/2103/2103.06312.pdf

[7] European Commission, Directorate-General for Communications Networks, Content and Technology, 'Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS'. CNECT, Apr. 21, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206

[8] L. Schmitt, 'Mapping global AI governance: a nascent regime in a fragmented landscape', *AI Ethics*, vol. 2, 2021, doi: 10.1007/s43681-021-00083-y.

[9] CEN CENELEC, 'CEN-CENELEC Focus Group Report - Road Map on Artificial Intelligence (AI)', CEN CENELEC, 2020. [Online]. Available: https://www.standict.eu/node/4854

[10] Laura Degallaix and Mariolina Eliantonio, 'THE USE OF STANDARDS IN LEGISLATION AND POLICIES: How to ensure a legally-sound and socially-acceptable use of standards by policy-makers in view of the CJEU James Elliott ruling and other court cases', ECOS - European Environmental Citizens Organisation for Standardisation, Jun. 2018. [Online]. Available: https://ecostandard.org/wp-content/uploads/2018-06-11-The-use-of-standards-in-legislation-and-policies-ECOS-discussion-paper.pdf

[11] International Organisation for Standardisation (ISO), 'BENEFITS OF STANDARDS: ISO AND POLICY MAKERS'. International Organisation for Standardisation (ISO). [Online].

Available: https://www.iso.org/iso-and-policy-makers.html
[12] G. Hepburn, 'Alternatives to traditional regulation', *Paris Organ. Econ. Coop. Dev.*, 2006.
[13] T. Zielke, 'Is Artificial Intelligence Ready for Standardization?', in *Commun. Comput. Info. Sci.*, 2020, vol. 1251 CCIS, pp. 259–274. doi: 10.1007/978-3-030-56441-4_19.
[14] I. Mezgár and J. Váncza, 'From ethics to standards – A path via responsible AI to cyber-physical production systems', *Annu. Rev. Control*, vol. 53, pp. 391–404, Jan. 2022, doi: 10.1016/j.arcontrol.2022.04.002.
[15] N. Brown *et al.*, 'The Role of Standardisation in Support of Emerging Technologies in the UK', Department for Business, Energy and Industrial Strategy (BEIS), London, May 2022. [Online]. Available: https://www.gov.uk/government/publications/role-of-standardisation-in-support-of-emerging-technologies
[16] The British Standards Institution (BSI), 'BSI White Paper – Overview of standardization landscape in artificial intelligence', The British Standards Institution (BSI), 2019. [Online]. Available: https://www.bsigroup.com/en-GB/about-bsi/media-centre/press-releases/2019/september/bsi-calls-for-collaboration-on-artificial-intelligence-guidance-to-give-industry-clarity/
[17] P. Cihon, 'Technical Report, Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development', *Future Humanity Inst. Univ. Oxf.*, 2019.
[18] T. Schmid, W. Hildesheim, T. Holoyad, and K. Schumacher, 'The AI Methods, Capabilities and Criticality Grid', *KI - Künstl. Intell.*, vol. 35, no. 3, pp. 425–440, Nov. 2021, doi: 10.1007/s13218-021-00736-4.
[19] International Organisation for Standardisation (ISO), 'STANDARDS BY ISO/IEC JTC 1/SC 42 Artificial intelligence', *STANDARDS BY ISO/IEC JTC 1/SC 42 Artificial intelligence*, 2022. https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0
[20] Institute of Electrical and Electronics Engineers (IEEE), 'IEEE portfolio of AIS technology and impact standards and standards projects', *IEEE portfolio of AIS technology and impact standards and standards projects*, 2022. https://standards.ieee.org/initiatives/artificial-intelligence-systems/standards/#p7000
[21] European Commission, Joint Research Centre, S. Nativi, and S. De Nigris, *AI Watch, AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework*. Publications Office, 2021. doi: 10.2760/376602.
[22] The British Standards Institution (BSI), 'BSI Standards Development', *ART/1 - Artificial Intelligence*, 2022. https://standardsdevelopment.bsigroup.com/committees/50281655#published
[23] B. Cheatham, K. Javanmardian, and H. Samandari, 'Confronting the risks of artificial intelligence', *McKinsey Q.*, vol. 2019, no. 2, 2019, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079283855&partnerID=40&md5=86169ee128e7c7f33dbf9659a97fda36
[24] A. Koene, G. Ezeani, R. Kumar, N. Santiago, and D. Wright, 'A survey of artificial intelligence risk assessment methodologies. The global state of play and leading practices

identified', Trilateral Research and Ernst & Young (EY), 2022. [Online]. Available: https://www.trilateralresearch.com/wp-content/uploads/2022/01/A-survey-of-AI-Risk-Assessment-Methodologies-full-report.pdf

[25] National Institute of Standards and Technology, 'AI Risk Management Framework: Initial Draft'. NIST, 2022. [Online]. Available: https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

[26] Mark McFadden, Kate Jones, Emily Taylor, and Georgia Osborn, 'Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation', Oxford Information Labs, Dec. 2021. [Online]. Available: https://oxil.uk/publications/2021-12-02-oxford-internet-institute-oxil-harmonising-ai/Harmonising-AI-OXIL.pdf

[27] A. Steimers and M. Schneider, 'Sources of Risk of AI Systems', *Int. J. Environ. Res. Public. Health*, vol. 19, no. 6, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126452843&doi=10.3390%2fijerph19063641&partnerID=40&md5=3c009655eaebb1c6f2fe1c7136b906e6

[28] National Institute of Standards and Technology (NIST), 'AI Risk Management Framework Concept Paper'. National Institute of Standards and Technology (NIST), Dec. 13, 2021. [Online]. Available: https://www.nist.gov/system/files/documents/2021/12/14/AI%20RMF%20Concept%20Paper_13Dec2021_posted.pdf

[29] P. B. de Laat, 'Companies Committed to Responsible AI: From Principles towards Implementation and Regulation?', *Philos. Technol.*, vol. 34, no. 4, pp. 1135–1193, 2021.

[30] Lilian Edwards, 'Regulating AI in Europe: four problems and four solutions', Ada Lovelace Institute, Expert Opinion, Mar. 2022. [Online]. Available: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf

[31] A. Luusua and J. Ylipulli, 'Artificial intelligence and risk in design', in *DIS Companion - Companion Publ. ACM Des. Interact. Syst. Conf.*, ["University of Oulu, Oulu, Finland", 'Aalto University, Espoo, Finland'], 2020, pp. 1235–1244. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090503450&doi=10.1145%2f3357236.3395491&partnerID=40&md5=e2165f582ad1f38921d57648ccd4da42

[32] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi, 'Operationalising AI ethics: barriers, enablers and next steps', *AI Soc.*, 2021, doi: 10.1007/s00146-021-01308-8.

[33] J. Mökander and L. Floridi, 'Operationalising AI governance through ethics-based auditing: an industry case study', *AI Ethics*, May 2022, doi: 10.1007/s43681-022-00171-7.

[34] L. Floridi, 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical', *Philos. Technol.*, vol. 32, no. 2, pp. 185–193, Jun. 2019, doi: 10.1007/s13347-019-00354-x.

[35] Tim McGarr, 'Unlocking the benefits of artificial intelligence with standards', May 2022. https://www.iec.ch/blog/unlocking-benefits-artificial-intelligence-standards

[36] Ellen Hohma, Auxane Boch, and Rainer Trauth, 'White Paper: Towards an Accountability Framework for Artificial Intelligence Systems', TUM IEAI and Fujitsu Limited, Aug. 2022. [Online]. Available: https://www.ieai.sot.tum.de/ieai-white-paper-series-towards-an-accountability-framework-for-artificial-intelligence-systems/

[37] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, 'From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices', *Sci. Eng. Ethics*, vol. 26, no. 4, pp. 2141–2168, Aug. 2020, doi: 10.1007/s11948-019-00165-5.

[38] M. Brundage *et al.*, 'Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims'. arXiv, 2020. doi: 10.48550/ARXIV.2004.07213.

[39] Laurie A. Cumbo, *Automated employment decision tools*. 2021. [Online]. Available: https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search

[40] National Institute of Standards and Technology (NIST), 'AI Measurement and Evaluation'. National Institute of Standards and Technology (NIST), May 16, 2022. [Online]. Available: https://www.nist.gov/programs-projects/ai-measurement-and-evaluation

[41] J. Krijger, 'Enter the metrics: critical theory and organizational operationalization of AI ethics', *AI Soc.*, 2021, doi: 10.1007/s00146-021-01256-3.

[42] N. A. Smuha, 'From a "race to AI" to a "race to AI regulation": regulatory competition for artificial intelligence', *Law Innov. Technol.*, vol. 13, no. 1, pp. 57–84, 2021.

[43] R. Clarke, 'Regulatory alternatives for AI', *Comput. Law Secur. Rev.*, vol. 35, no. 4, pp. 398–409, Aug. 2019, doi: 10.1016/j.clsr.2019.04.008.

[44] Ilina Georgieva, Tjerk Timan, and Marissa Hoekstra, 'Regulatory divergences in the draft AI act: Differences in public and private sector obligations'. European Parliamentary Research Service, Scientific Foresight Unit (STOA), May 2022. [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729507/EPRS_STU(2022)729507_EN.pdf

[45] Russell T. Vought, 'MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: Guidance for Regulation of Artificial Intelligence Applications'. 2019. [Online]. Available: https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

[46] Sotirios Dimopoulos and Mattia Ferrini, 'The future of the AI regulatory landscape: are you ready?', *KPMG Switzerland Blog: The future of the AI regulatory landscape: are you ready?*, Jun. 25, 2021. https://home.kpmg/ch/en/blogs/home/posts/2021/06/ai-regulatory-landscape.html

[47] Chanley T. Howell, 'AI Regulation: Where do China, the EU, and the U.S. Stand Today?' The National Law Review, Aug. 03, 2022. [Online]. Available: https://www.natlawreview.com/article/ai-regulation-where-do-china-eu-and-us-stand-today

[48] Rt Hon Nadine Dorries MP, 'Policy paper Establishing a pro-innovation approach to regulating AI', Digital, Culture, Media and Sport (DCMS), London, Policy Paper, Jul. 2022. [Online]. Available: https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-

policy-statement
[49] Stephan Zimprich, 'The EU-AI Regulation - Part 1: Overview and structure'. Fieldfisher, Aug. 24, 2021. [Online]. Available: https://www.fieldfisher.com/en/insights/the-eu-ai-regulation-part-1
[50] Rogier Creemers, Graham Webster, and Helen Toner, 'Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022'. Stanford University, DIGICHINA, Jan. 10, 2022. [Online]. Available: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/
[51] Daniel Leufer, 'Feedback from: Access Now Europe'. European Commission, Aug. 05, 2021. [Online]. Available: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665462_en
[52] L. A. Bygrave, 'The "Strasbourg Effect" on data protection in light of the "Brussels Effect": Logic, mechanics and prospects', *Comput. Law Secur. Rev.*, vol. 40, p. 105460, Apr. 2021, doi: 10.1016/j.clsr.2020.105460.
[53] Helen Toner, Paul Triolo, and Rogier Creemers, 'Experts Examine China's Pioneering Draft Algorithm Regulations'. Stanford University, DIGICHINA, Aug. 27, 2021. [Online]. Available: https://digichina.stanford.edu/work/experts-examine-chinas-pioneering-draft-algorithm-regulations/
[54] E. Kazim *et al.*, 'Innovation and opportunity: review of the UK's national AI strategy', *Discov. Artif. Intell.*, vol. 1, Dec. 2021, doi: 10.1007/s44163-021-00014-0.
[55] Nicol Turner Lee and Samantha Lai, 'The U.S. can improve its AI governance strategy by addressing online biases', May 17, 2022. https://www.brookings.edu/blog/techtank/2022/05/17/the-u-s-can-improve-its-ai-governance-strategy-by-addressing-online-biases/
[56] National Institute of Standards and Technology (NIST), 'A Plan for Federal Engagement in Developing AI Technical Standards and Related Tools in response to Executive Order (EO 13859)'. National Institute of Standards and Technology (NIST), Apr. 05, 2022. [Online]. Available: https://www.nist.gov/artificial-intelligence/plan-federal-engagement-developing-ai-technical-standards-and-related-tools
[57] P. Almeida, C. Santos Jr, and J. Farias, 'Artificial Intelligence Regulation: a framework for governance', *Ethics Inf. Technol.*, vol. 23, Sep. 2021, doi: 10.1007/s10676-021-09593-z.
[58] K. Jia and N. Zhang, 'Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines', *Electron. Mark.*, vol. 32, no. 1, pp. 59–71, 2022.
[59] Ed Van Buren, Bruce Chew, and William D. Eggers, 'AI Readiness for Government: Are you ready for AI? As various government agencies prepare to deploy artificial intelligence, a six-pronged framework can help them assess their AI readiness'. Deloitte Centre for Government Insights, Jan. 24, 2020. [Online]. Available: https://www2.deloitte.com/za/en/pages/public-sector/articles/AI-Readiness-for-Government.html
[60] Lilian Edwards, 'Expert explainer: The EU AI Act Proposal'. Ada Lovelace Institute, Apr. 2022. [Online]. Available: https://www.adalovelaceinstitute.org/wp-

content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf
[61] European Institute of Public Administration (EIPA), 'The Artificial Intelligence Act Proposal and its Implications for Member States'. European Institute of Public Administration (EIPA), Sep. 2021. [Online]. Available: https://www.eipa.eu/publications/briefing/the-artificial-intelligence-act-proposal-and-its-implications-for-member-states/
[62] B. C. Stahl, R. Rodrigues, N. Santiago, and K. Macnish, 'A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values', *Comput. Law Secur. Rev.*, vol. 45, p. 105661, Jul. 2022, doi: 10.1016/j.clsr.2022.105661.
[63] Ada Lovelace Institute, 'Regulate to innovate: A route to regulation that reflects the ambition of the UK AI Strategy'. Ada Lovelace Institute, Nov. 2021. [Online]. Available: https://www.adalovelaceinstitute.org/report/regulate-innovate/
[64] Department for Digital, Culture, Media and Sport (DCMS) and Rt Hon Damian Collins MP, 'Press release: UK sets out proposals for new AI rulebook to unleash innovation and boost public trust in the technology'. Department for Digital, Culture, Media and Sport (DCMS), Jul. 18, 2022. [Online]. Available: https://www.gov.uk/government/news/uk-sets-out-proposals-for-new-ai-rulebook-to-unleash-innovation-and-boost-public-trust-in-the-technology
[65] Peter Waters, 'Are regulators "Fit for AI"?' Lexology, Aug. 16, 2022. [Online]. Available: https://www.lexology.com/library/detail.aspx?g=65bb0347-fc24-4541-88f5-d8645523a97d
[66] J. Ibanez and M. Olmeda, 'Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study', *AI Soc.*, doi: 10.1007/s00146-021-01267-0.
[67] Organisation for Economic Co-operation and Development (OECD), 'OECD AI Principles overview', 2022. https://oecd.ai/en/ai-principles
[68] Microsoft, 'Responsible AI', 2022. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6
[69] European Commission, 'Ethics guidelines for trustworthy AI', European Commission, Apr. 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[70] Future of Life Institute, 'ASILOMAR AI PRINCIPLES', 2017. https://futureoflife.org/2017/08/11/ai-principles/
[71] B. Rakova, J. Yang, H. Cramer, and R. Chowdhury, 'Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for shifting Organizational Practices', *CoRR*, vol. abs/2006.12358, 2020, [Online]. Available: https://arxiv.org/abs/2006.12358
[72] B. Beckert, 'The European way of doing Artificial Intelligence: The state of play implementing Trustworthy AI', in *2021 60th FITCE Communication Days Congress for ICT Professionals: Industrial Data – Cloud, Low Latency and Privacy (FITCE)*, Sep. 2021, pp. 1–8. doi: 10.1109/FITCE53297.2021.9588560.
[73] A. Jobin, M. Ienca, and E. Vayena, 'The global landscape of AI ethics guidelines', *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.
[74] Joshua P. Meltzer, 'The impact of artificial intelligence on international trade', Brookings Institution, Dec. 2018. [Online]. Available: https://www.brookings.edu/research/the-impact-

of-artificial-intelligence-on-international-trade/#:~:text=As%20AI%20is%20based%20on,to%20respond%20to%20this%20concern.
[75] M. S. Jawad, H. Mahdin, N. A. Mohammed Alduais, M. Hlayel, S. A. Mostafa, and M. H. Abd Wahab, 'Recent and Future Innovative Artificial Intelligence Services and Fields', in *2021 4th International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR)*, Sep. 2021, pp. 29–32. doi: 10.1109/ISAMSR53229.2021.9567891.
[76] Beena Ammanath, Susanne Hupfer, and David Jarvis, 'Thriving in the era of pervasive AI Deloitte's State of AI in the Enterprise, 3rd Edition', Deloitte AI Institute and the Deloitte Centre for Technology, Media and Telecommunications, 2020.
[77] International Business Machines (IBM), 'IBM Global AI Adoption Index 2022', International Business Machines (IBM), 2022. [Online]. Available: https://www.ibm.com/watson/resources/ai-adoption
[78] Roland Berger, 'AI STARTUPS AS INNOVATION DRIVERS', Roland Berger, May 2018. [Online]. Available: https://www.rolandberger.com/en/Insights/Publications/AI-startups-as-innovation-drivers.html
[79] M. Rimol, 'Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation', Gartner, Sep. 2021. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2021-09-07-gartner-identifies-four-trends-driving-near-term-artificial-intelligence-innovation
[80] R. Fulton, D. Fulton, and S. Kaplan, 'Artificial Intelligence: Framework of driving triggers to past, present and future applications and influencers of industry sector adoption'. arXiv, 2022. doi: 10.48550/ARXIV.2204.01518.
[81] Y. K. Dwivedi *et al.*, 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy', *Int. J. Inf. Manag.*, vol. 57, p. 101994, Apr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.
[82] G. Cecere, N. Corrocher, C. Gossart, and M. Ozman, 'Lock-in and path dependence: an evolutionary approach to eco-innovations', *J. Evol. Econ.*, vol. 24, no. 5, pp. 1037–1065, Nov. 2014, doi: 10.1007/s00191-014-0381-5.
[83] Andrew Evans and Anja Heimann, 'AI ACTIVITY IN UK BUSINESSES: An assessment of the scale of AI activity in UK businesses and scenarios for growth over the next twenty years', Department for Digital, Culture, Media, and Sport (DCMS), Jan. 2022. [Online]. Available: https://www.gov.uk/government/publications/ai-activity-in-uk-businesses
[84] Department for Digital, Culture, Media and Sport (DCMS), 'Data foundations and AI adoption in the UK private and third sectors: Executive Summary'. Department for Digital, Culture, Media and Sport (DCMS), Aug. 16, 2021. [Online]. Available: https://www.gov.uk/government/publications/data-foundations-and-ai-adoption-in-the-uk-private-and-third-sectors/data-foundations-and-ai-adoption-in-the-uk-private-and-third-sectors-executive-summary
[85] J. Clark and G. K. Hadfield, 'Regulatory Markets for AI Safety'. arXiv, 2020. doi: 10.48550/ARXIV.2001.00078.
[86] T. J. Foxon, 'Technological and institutional " lock-in " as a barrier to sustainable

innovation', 2002. [Online]. Available: https://www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/icept/7294726.PDF

[87] European Commission, 'COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS An EU Strategy on Standardisation Setting global standards in support of a resilient, green and digital EU single market'. European Commission, Feb. 02, 2022. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0031

[88] E. Barrance, E. Kazim, A. Hilliard, M. Trengove, S. Zannone, and A. Koshiyama, 'Overview and commentary of the CDEI's extended roadmap to an effective AI assurance ecosystem', *Front. Artif. Intell.*, vol. 5, 2022, [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2022.932358

[89] International Business Machines (IBM), 'The new AI innovation equation'. International Business Machines (IBM). [Online]. Available: https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-innovation-equation.html#section1

[90] David Lucatch, 'What's In Store For Web 3.0 In 2022', Forbes, Feb. 2022. [Online]. Available: https://www.forbes.com/sites/forbesbusinesscouncil/2022/02/22/whats-in-store-for-web-30-in-2022/?sh=3bc27c166a07

[91] Jon Henley, 'Chess robot grabs and breaks finger of seven-year-old opponent: Moscow incident occurred because child "violated" safety rules by taking turn too quickly, says official', The Guardian, Jul. 2022. [Online]. Available: https://www.theguardian.com/sport/2022/jul/24/chess-robot-grabs-and-breaks-finger-of-seven-year-old-opponent-moscow

[92] Lance Elliot, 'AI Ethics And That Viral Story Of The Chess Playing Robot That Broke The Finger Of A Seven-Year-Old During A Heated Chess Match Proffers Spellbinding Autonomous Systems Lessons', Forbes, Jul. 2022. [Online]. Available: https://www.forbes.com/sites/lanceeliot/2022/07/27/ai-ethics-and-that-viral-story-of-the-chess-playing-robot-that-broke-the-finger-of-a-seven-year-old-during-a-heated-chess-match-proffers-spellbinding-autonomous-systems-lessons/?sh=1106e1f65b2b

[93] L. E. McCray, K. A. Oye, and A. C. Petersen, 'Planned adaptation in risk regulation: An initial survey of US environmental, health, and safety regulation', *Two Spec. Sect. Risk Technol. Addressing Deep Uncertain. Using Adapt. Policies*, vol. 77, no. 6, pp. 951–959, Jul. 2010, doi: 10.1016/j.techfore.2009.12.001.

[94] Government Office for Science, 'The Futures Toolkit: Tools for Futures Thinking and Foresight Across UK Government'. Government Office for Science, Nov. 2017. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674209/futures-toolkit-edition-1.pdf

ANNEXES

ANNEX A – Methodology

1. Literature review protocol(s)

The literature review was conducted using a systematic-style approach. Before conducting the review, and considering the broad scope of this research, a protocol for each cluster (AI risks and harms, AI regulation, policy, and strategy, and AI and innovation) was created. Nevertheless, standardised inclusion and exclusion criteria were identified for all three clusters. The role standards play in AI governance was mainly incorporated into the regulation, policy, and strategy section. However, when reviewing the literature, all clusters contributed to collecting evidence concerning the role of standards.

Protocols inclusion and exclusion criteria

Inclusion criteria	• Publications between 2018 and 2022 • Studies with all types of research design and methodologies • Prioritisation of articles/documents providing an "overview" of cluster topics • Academic papers, policy documents (including white papers and regulations), and grey literature (reports from consulting organisations) • Specific articles outside of criteria were retroactively added where extremely valuable
Exclusion criteria	• Publications which focus on only one specific application or technique, or are not generalisable to other domains • Medicine and health-related publications • Publications not available in English • Retracted publications

Protocol: AI risks and harms

Research questions	1. What are the major risks associated with the development and use of AI? 2. What are the challenges of assessing and mitigating AI risks and harms?

Search strategy	**Database** 1. Scopus (Academic Literature) 2. IEEE Xplore Digital Library (Academic Literature) 3. Harvard Kennedy School Library and Knowledge Service (Grey Literature) **Syntax** 1. Scopus: *Title, Abstract Level: Risks and Harms Keywords* *("Risk" W/5 ai) OR ("Harm" W/5 ai) OR ("Risk" W/5 "Artificial Intelligence") OR ("Harm" W/5 "Artificial Intelligence")* *Title level: Artificial Intelligence Keywords* *"AI" OR "Artificial intelligence"* 2. IEEE Xplore Digital Library: *Title level: Artificial Intelligence Keywords* *"AI" OR "Artificial intelligence"* *Title, Abstract Level: Risks and Harms Keywords* *"Risk" OR "Harm"* 3. Harvard Kennedy School, Library and Knowledge *Generic Search: Artificial Intelligence, Risks, and Standards Keywords* *"Artificial Intelligence Risk" and Standard*

Protocol: AI regulation, policy and strategy

Research questions and sub-questions	1. What is the current landscape of AI regulation? a. Sub-question: What are the emerging trends in AI regulation? b. Sub-question: How are countries (UK, USA, EU and China) approaching AI regulation? 2. What is the role of standards in AI regulation? a. Sub-question: Are there gaps in AI regulation that

	standards can complement?
Search syntax	**Database**: 1. Scopus (Academic Literature) 2. IEEE Xplore Digital Library (Academic Literature) 3. Google academic (Grey Literature) **Syntax** ***Scopus, IEEE:*** ● *Syntax 1a* ○ *"artificial intelligence" AND/OR "AI"* ○ *AND "Regulat*" OR "law" OR "legislat*" AND/OR "Polic*" and/or ""GOVERN*"* ○ *AND UK or/and US or/AND EU and/China* ● *Syntax 1b* ○ *"artificial intelligence" AND/OR "AI"* ○ *AND "Regulat*" OR "law" OR "legislati*" AND/OR "Polic*"* ● *Syntax 2* ○ *"artificial intelligence" AND/OR "AI"* ○ *AND "Regulat*" OR "law" OR "legislati*" OR "Polic*"* ○ *AND "Standards" OR "standardisation" OR "standardization"* ***Google and Google Scholar (early step to access grey lit)*** ● *"AI regulation" (EU or USA or UK or China)* ● *"AI regulation' or "AI governance" or "AI policy"* ● *"AI standards" and "regulation or policy"*

Protocol: AI innovation

Research questions	1. What are the key drivers of AI innovation? 2. What are the key barriers to AI innovation?

Search strategy	**Database**: 1. Scopus (Academic Literature) 2. IEEE Xplore Digital Library (Academic Literature) 3. Harvard Kennedy School, Library and Knowledge Service (Grey Literature) 4. Google Scholar (Grey Literature) **Syntax: What are the key drivers of AI innovation?** ***Scopus*** ● TITLE-ABS-KEY ((ai OR "artificial intelligence") AND (innovat* OR develop* OR "cutting edge" OR novel) AND (encourage OR promote OR facilitate OR foster *OR trigger*)) AND PUBYEAR > 2018 *Exclude Medicine* ***IEEE*** ● ((AI OR artificial intelligence") AND (innovat* OR develop* OR novel) AND (trigger OR promote OR drive* OR foster OR encourage)) >2018 ● "artificial intelligence" innovation >2018 ● ("All Metadata":AI) AND ("All Metadata":innovat*) AND ("All Metadata":(drive* OR "All Metadata":trigger OR "All Metadata":promote OR "All Metadata":foster)) ● ("All Metadata":AI) AND ("All Metadata":standards) AND ("All Metadata":innova*), date range 2019 – 2022 ***Harvard Kennedy School Think Tank Search*** "AI innovation", checked sort by both Date and Relevance "Drivers of AI innovation", checked sort by both Date and Relevance ***Open Think Tank Directory*** Topics: "Governance/Transparency", "Private Sector Development", and "Technology/Innovation"

2. Semi-structured interviews

Interview process and protocol

A total of 21 stakeholders were interviewed between 28th June and 2nd August 2022. All interviews were conducted virtually using Microsoft Teams, following UCL guidelines. Participants were asked to complete a consent form prior to the interview as per the Ethics Application. Two team members conducted each interview. The interviews were recorded for transcription purposes. All interviews were conducted within the UK.

Interviews served three goals: 1) validate or counter findings from the literature review, 2) gather in-depth insights from participants' experiences and 3) identify potential drivers of change for the future of the AI standardisation landscape. Interviews were semi-structured to allow for follow-up insights and explanation. A list of sample questions containing the main questions for each cluster is presented at the end of this section.

We analysed interviews using thematic analysis, and used an inductive approach to identify and examine key patterns and themes within data.

Participant backgrounds

The following table provides more information on stakeholders and their interviews. In total, we interviewed ten participants from industry, six from academia, two from government and three from SDOs. Participants' insights were cited in the report using the participant code below (as "P" and an identifying number, e.g., P12).

Participant Code	Stakeholder Category	Date (2022)
P1	Industry	28 June
P2	Industry	28 June
P3	Industry	1 July
P4	Academia	5 July
P5	Industry	12 July
P6	Industry	13 July
P7	Government	15 July
P8	Academia	18 July
P9	Industry	27 July
P10	Industry	26 July
P11	Government	21 July
P12	Industry	25 July
P13	Academia	28 July
P14	Standards Body	20 July

P15	Standards Body	28 July
P16	Standards Body	8 August
P17	Academia	11 August
P18	Industry	22 July
P19	Academia	4 August
P20	Industry	1 August
P21	Academia	2 August

Table 1: *Participant list by stakeholder category*

Sample interview questions

Introduction question:

- What led you to become interested in AI? What is the focus of your current work or research?

AI and regulation/policy/strategy

- Do you think there is a need for AI regulation?
- Do you have any concerns about the regulations that are currently being proposed? (e.g., impact on organisations, change in expectations for compliance, burdens, etc.)

AI risks and harms

- What are the potential risks emerging from AI development and use that concern your organisation most?
- Which risk category do you think needs the most attention from standards-makers? (e.g., explainability, bias, safety)?

AI and innovation

- What are the barriers that organisations are/your organisation is facing in adopting or developing AI?
- Can standards help find a balance between AI innovation and potential risks? If so, how?

AI governance and standards:

- What elements would AI standards need to support within current or upcoming regulatory proposals?
- What role could standards play in anticipating and mitigating AI-related risks and harms?

- What would you want to see in AI standards that would help support your organisation, especially in AI innovation?
- As AI evolves over the next three to five years, what role do you think standards should play?

3. Workshop

A workshop on the future of AI standards was conducted with standards-makers. This workshop was conducted via Microsoft Teams and used Mural, a virtual workshop web tool. It was conducted on 26th July 2022 and with a duration of one hour.

The workshop served two purposes: 1) identifying and exploring drivers of change in the AI standardisation landscape according to their impact and anticipated timeline; and 2) eliciting future needs of effective AI governance and AI standards more specifically.

The workshop Mural board can be viewed by following this link: https://bit.ly/3pEtLIR

Workshop participants

In total, nine participants attended the workshop. All participants are involved in AI standardisation efforts. Their stakeholder representation can be found below in Table 2.

Stakeholder Category	Number of Participants
Industry	5
Standards body	1
Academia	2
Government	1

Table 2: *Workshop participants by stakeholder category*

Workshop activities

To achieve the aforementioned goals, two activities were designed.

Activity 1: During this activity, experts were asked to brainstorm drivers of change which might impact standardisation needs. We asked participants to focus on near-term drivers. As ***Figure 1*** shows, technological, regulatory, financial and other categories were provided. Examples were given

in advance to encourage participants to generate ideas. After brainstorming drivers of change, participants were asked to assign a likelihood of occurrence between one and ten years, which is also displayed in ***Figure 1***.

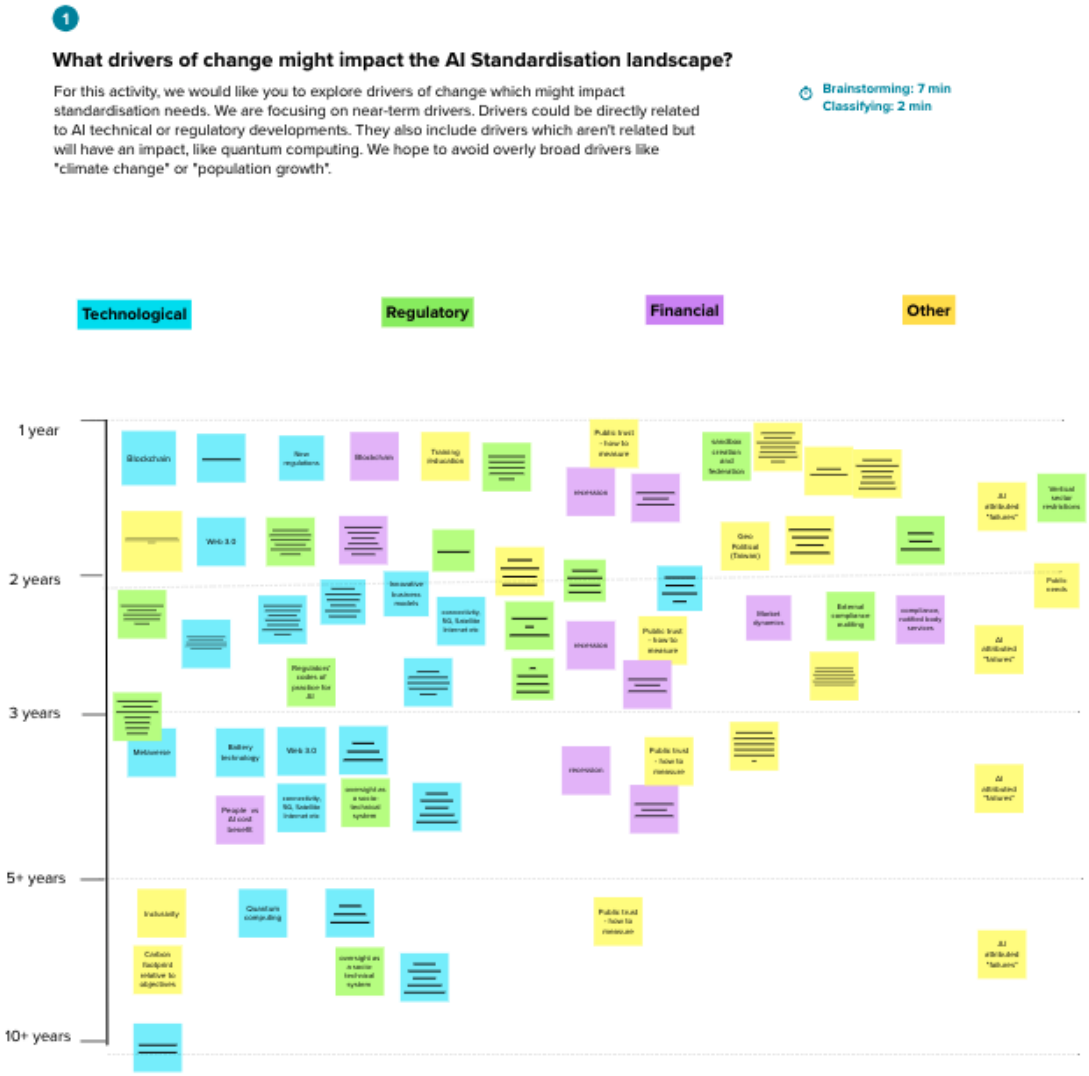


Figure 1: *Workshop activity 1*

Activity 2: Activity 2 consisted of two parts. First, participants were asked to brainstorm future AI governance needs. The goal of the workshop design was to prime participants to generate drivers of change in the first activity that would help them imagine future needs they might not have otherwise considered. Four categories were provided, with examples from each: regulatory/legal needs; innovation needs; safety and trustworthiness needs; and organisational and operational needs.

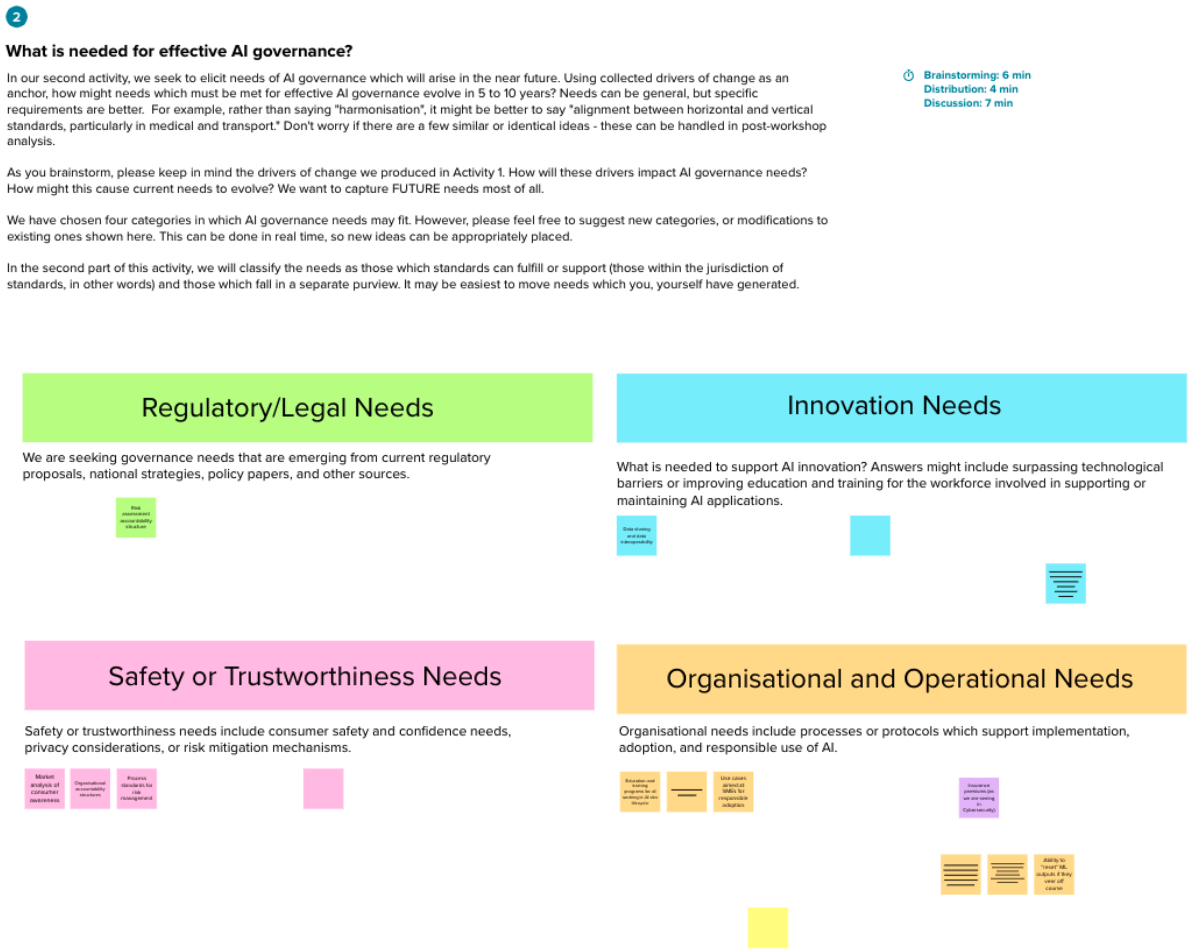


Figure 2: *Workshop Activity 2 (Part 1)*

Second, as showcased in Figure 3, participants were given a categorisation exercise to classify the needs by those which standards can fulfil or support and those which fall under a different purview.

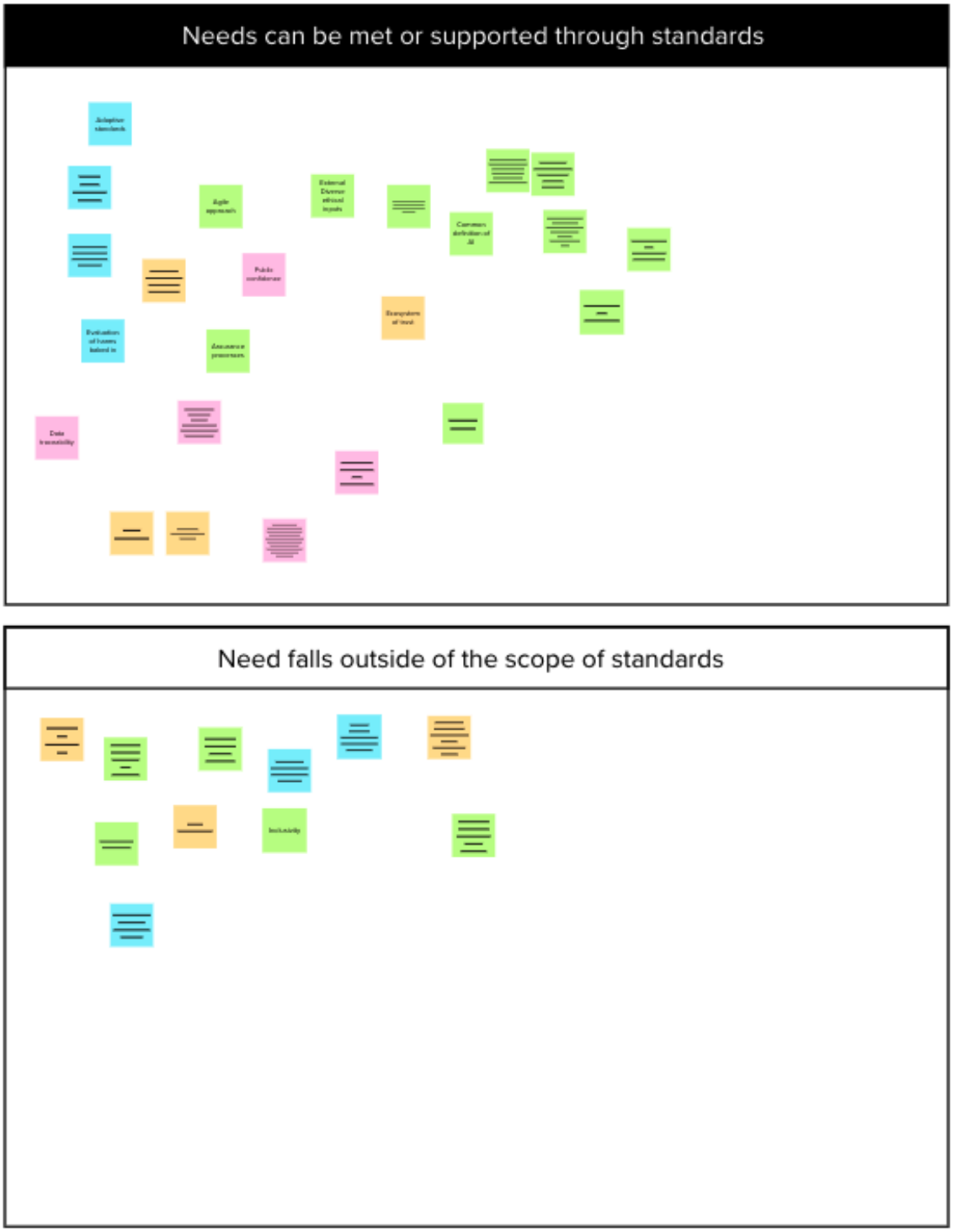


Figure 3: *Workshop activity 2 (Part 2)*

Final discussion: Following these activities, two questions were put forth for discussion:

- Which needs should be prioritised by standards bodies in the next three to five years and why?; and
- Are there any concepts generated within this workshop that, in your opinion, paint an inaccurate picture? Is anything missing?

Workshop output

Insights from this workshop were then used to solidify and prioritise a list of drivers of change (see ***Annex D***), incorporate and validate AI governance needs from interviews and the literature, and validate which future needs would be well suited to standards.

4. Horizon scanning

A horizon scanning activity was conducted to ensure that the collected standardisation needs were both future-facing and informed by relevant drivers of change in the field of AI. For that purpose, this activity focused on gathering and generating insights on drivers of change affecting the landscape of AI governance in the next one to ten years. This activity was conducted throughout the research project.

While a horizon scanning activity tends to focus on scanning literature and conducting workshop discussion [1], this research has broadened the scope of scanned information to include semi-structured interviews. This was done to develop a more robust evidence base while enriching the gathered drivers of change with expert and stakeholder input.

Steps

In conducting horizon scanning activities, we:

1. Gathered and generated drivers of change from literature reviews, semi-structured interviews, and a workshop via discussion by experts and stakeholders
2. Collected an initial list of 106 drivers of change
3. Held an internal discussion to provide additional context for certain drivers of change that did not have a description or context provided by the stakeholder involved in the workshop
4. Reduced the drivers of change by eliminating redundancy, obscurity, (i.e., 'economic conditions' without appropriate description or context), and combined drivers of change applied to multiple contexts (i.e., Blockchain, which was included in the Financial and Other categories in the workshop)
5. Finalised a list of 47 drivers of change
6. Analysed and plotted the drivers of change according to their source of origin and anticipated time period between one to ten years

7. Established themes according to the gathered drivers of change, translating them into categories: Regulatory/Governmental, Technological, Socio-Technical, Data-Driven, and Economic
8. Conducted further analysis on drivers of change (see ***Annex D*** and ***3.6.***)

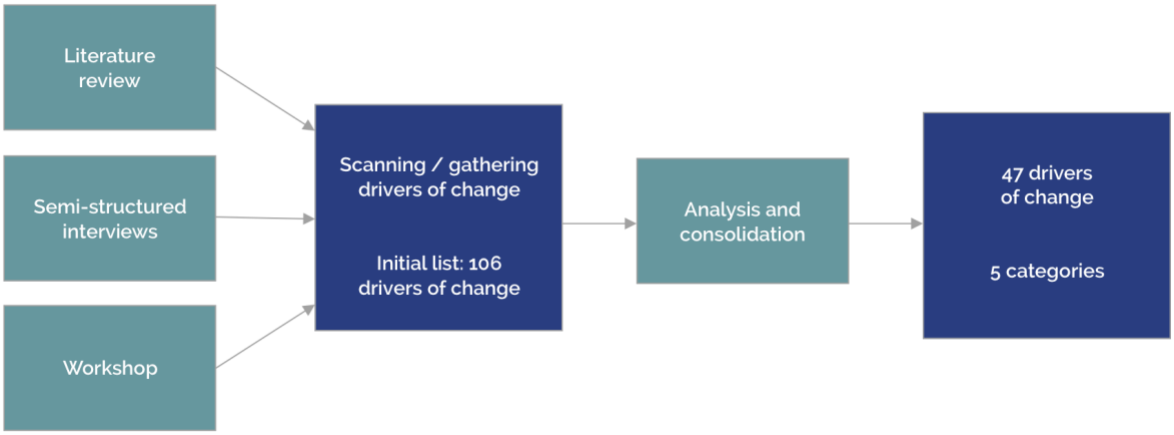


Figure 4: *Horizon scanning activity*

5. Ethics application

Our research was carried out in accordance with the ethics guidelines set out by the UCL Research Ethics Committee, and all data protection legislation (GDPR and the Data Protection Act 2018). We declared all data collection and storage methods, as well as our precautionary measures to mitigate any risks associated with the project during our ethics application process. We also requested all participants of the research to provide written consent before all data collection processes to ensure all information is consensually given.

6. Reference

[1] Government Office for Science. "The Futures Toolkit: Tools for Futures Thinking and Foresight Across UK Government." Government Office for Science, November 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674209/futures-toolkit-edition-1.pdf.

ANNEX B - Literature review findings

Introduction

While research on Artificial Intelligence (AI) governance is growing rapidly [1], academic and grey literature regarding the role of standards in AI governance is still limited. With this in mind, we conducted a literature review to gather a comprehensive understanding of key debates, challenges, and conceptual issues around the role of standards in addressing emerging industry and regulatory needs. To do so, we followed the cluster-focus structure of this research. We divided this literature review into AI and innovation; AI regulation, policy, and strategy; and AI risks and harms. We also included an exploration of existing and emerging AI standards.

Aware of the broad nature of this task, we conducted the literature review using a systematic-style approach, following well-established processes and protocols commonly used in Rapid Evidence Assessments. For each cluster, both academic and grey literature sources were included to diversify the types of evidence used and provide a holistic picture of the topic. Furthermore, to ensure robustness, transparency and reproducibility, we developed a protocol with research questions, search strategy, syntax, database choices and inclusion and exclusion criteria. Further details of these protocols were presented in the previous section.

The following literature review is structured as follows. Firstly, we present the main findings related to AI risks and harms. Secondly, we introduce the AI regulation and policy landscape with key findings related to the countries leading the AI regulation race. Thirdly, we explore the drivers and barriers to AI adoption and innovation. Lastly, we included a section on AI standards' role, which condensed some of the clusters' findings.

AI risks and harms

In exploring AI risks and harms, this literature review aims to provide a systematic review and analysis of risks associated with AI development and use and document the most pressing challenges encountered with AI risk management.

The research questions were:

- What are the major risks associated with the development and use of AI?
- What are the challenges of assessing and mitigating AI risks and harms?

Despite many consumer benefits and considerable business value generated by AI, its increased adoption has raised several significant risks, some of which may lead to systemic discrimination, privacy violations, accidents, and deprivation of human autonomy [2]. Such failures, particularly in critical applications such as healthcare, public services, or military, will pose significant threats to human rights and public safety and will hinder AI adoption and acceptance around the world [3], [4]. Both academic and industry experts agree on the need to identify, mitigate, and minimise these risks to realise the full potential of AI technologies. This is in addition to ensuring the safety, trustworthiness, and beneficial aspects are maintained for the collective benefit [3], [5]. Within this context, the literature supports risk-based governance of AI, centring around the need to categorise and evaluate different types of AI risks.

What are the major risks associated with the development and use of AI?

This section presents a panoramic and non-exhaustive overview of the most prevalent AI risks covered in the literature. Although an in-depth analysis of each area of risk is outside of the scope of this review, the main purpose is to identify key gaps and associated mitigation challenges in risk-based AI governance. Evidence gathered from this review will enable identification of current and emerging standardisation needs to address the governance challenges uncovered.

Six major sources of risks were derived from the literature: (1) transparency, (2) privacy and data protection, (3) security, (4) safety, (5) bias, unfairness, and discrimination, (6) accountability.

Transparency

Many scholars identified the lack of AI system transparency as a significant issue, leading to information imbalance between AI developers, operators, and consumers [5], [6]. This characteristic is particularly important for effective risk mitigation, as transparency is considered as a necessary prerequisite for identifying and reducing AI risks. Terminologies around the risk dimension varied, with many publications interchangeably using transparency, explainability and interpretability to refer to the 'black box' nature of AI systems. However, standards bodies like NIST have urged to

clarify the distinction between explainability (how easily understood the inner workings of an AI system is) and interpretability (how easily model output can be understood and contextualised) [5].

The desire to enforce transparency into the design and performance of an AI system has proven to be a difficult endeavour, primarily due to the ever-increasing complexity of many AI systems on the market. Such complexity hinders the effort to explain the decision route that led an AI system to a particular output [7]. While many AI governance frameworks and ethical guidelines necessitate interpretation of AI systems' inner workings in nontechnical language, many studies highlight that even algorithm designers find it difficult to provide a clear explanation [8]. Additionally, even if machine learning decisions are explained, their decision-subjects may not necessarily agree with the outcome [6]. Without appropriate redress mechanisms in place, such explanations may not serve to protect consumers.

Privacy and data protection

AI risks associated with privacy and data protection undermine the ability of individuals to control or influence what information was inferred from them, how it may be collected and stored, and by whom that information may be disclosed [9], [10]. Many studies highlighted that AI systems render individuals with little control and oversight on how their personal data is appropriated and what inferences are made from them [6].

In the EU, the GDPR Articles 15, 16, 17, and 21 were commonly cited to be the obvious legal remedies in relation to AI privacy and data protection risks. However, as scholars have noted some aspect of the GDPR, such as Article 9 (Processing of Special Categories of Personal Data) and Article 22(3) (Automated Individual Decision-Making, Including Profiling), fall short in providing remedies to contest inferences or important decisions that were derived from them [6], [11]. This is especially critical in 'high-risk inferences' [6] that occupies discourses on accountability of automated decision-making processes in the public sector.

Security

Complex AI models, such as neural networks, are found to exhibit unique weaknesses that require a higher level of scrutiny [10]. While many present-day cyber threats are easily repelled by adopting established best practices or tested IT security protocols, the same cannot be said for many AI systems. Their novel and transboundary nature has made implementation of reliable security measures for vulnerable AI systems difficult to achieve. Prominently, given the opaque nature of AI systems, investigation of a security incident may lead to few or no visible traces to the source of the breach or attack [7].

In this space, experts have proposed numerous solutions to addressing security issues relevant to AI. These include (1) placing good protection and recovery mechanisms, (2) considering and addressing vulnerabilities in the design process, (3) engaging human analysts in critical decision-making, (4) utilising risk management programmes, and (5) software upgrades [6]. However, similar to the classic dilemma of IT security risk, effective response to AI security issues requires a proactive

and responsive use of cyber security tools, policies and mechanisms by both developers and users at all stages [6], which is often not the case in real life. This points to a significant challenge, especially in safety critical applications such as healthcare and transportation.

Safety

AI safety risk is often discussed in the context of physical interactions with AI systems in factories or on the roads, with autonomous vehicles (AV) representing the most urgent area requiring serious considerations for risk management [12]. However, AI safety concerns extend beyond physical harm and cover negative impacts on property and environment, although these areas are not extensively explored within literature [9]. Addressing environmental risks of AI systems is especially crucial, given their monumental computing resources and energy requirements [7], [9].

Bias, unfairness, and discrimination

While bias is not necessarily a negative phenomenon, it becomes problematic when it leads to unfair treatment of an individual or group of individuals, potentially resulting in systemic discrimination [8]. Given the significance and potential severity of this risk factor, various guidance documents and frameworks have been produced advising AI practitioners to perform regular assessments on the representativeness of their datasets or employ human-in-the-loop and 'open' algorithm approaches [6].

However, studies show that these tools and guidance documents are insufficient in many areas [13]. For example, 'open' algorithm approach does not necessarily translate to being easily understood by people and it may indirectly lead to more issues with regards to discoverability of previously private data [6]. Furthermore, studies point to the challenges of translating a complex normative principle like fairness into a specific technological design. While various formal definitions of fairness have been put forward, they are mutually exclusive technological interpretations and only a single definition can be satisfied at one time [14].

Accountability

Accountability for AI systems necessitates putting in place mechanisms that assign responsibility for trustworthy AI development, deployment and use [15]. In the absence of standardised risk management frameworks and clear guidance for assigning liability when risky outcomes are realised, scholars stress the 'accountability gap' is widening [6]. Within organisational settings, studies show that AI practitioners are not clear on who is ultimately responsible for ensuring alignment between ethical principles and product design, which significantly hinders responsible AI practices [16].

What are the challenges of assessing and mitigating AI risks and harms?

Lack of common language and consensus

Despite growing interest in risk-based governance approaches, the literature review revealed a fragmented and incomplete body of knowledge on AI risks. Crucially, there is a lack of common language when it comes to AI risks, from how risk itself is defined [10] to how AI risks are conceptualised and classified [17]. The terminologies used around AI risks are often incoherent and fragmented between different publishing fields. Additionally, most studies focus on specific risks based on case studies and fail to provide a comprehensive risk analysis and categorisation [3], [8]. Although recent approaches have begun producing risk taxonomies and management frameworks, most papers only describe AI risks superficially and do not provide a thorough understanding of AI risks and their implications [3], [10].

Challenges with AI risk assessment

AI systems give rise to unique risks. Although risk management practices are well established in many disciplines, they are limited in their ability to assess and mitigate AI specific risks [18]. AI risk management needs to account for new sources of risk that are unique to AI systems and assess novel elements of well-established sources of risk (e.g., safety). For example, opacity of AI systems is a risk factor that is specific and unique to AI, as it is a direct result of AI model complexity [18]. Often referred to as the 'black box' problem, the primary risk here is that AI system decisions can no longer be retraced or understood by humans due to their inscrutable nature [3].

Challenges with AI risk measurement

As AI risks are not well-defined or understood, they are hard to measure [8]. This is further complicated by the temporal dimension of AI risks and long and complicated supply chains involved in developing a single AI system [19]. Furthermore, the opaque nature of AI systems makes auditability and traceability of AI systems especially challenging [20].

AI risks are contextual

At its core, risk is a human concept that is shaped by our values [21]. What we consider to be 'at risk' is guided by what we deem valuable and worth protecting. As a result, significant cultural differences are starting to emerge in the literature around AI risks. For example, the EU is considering banning live facial recognition technologies (FRTs) due to their significant impact on citizens' right to privacy [21]. However, in China, FRTs are widely used by the government to observe citizens in public spaces as illegal or antisocial behaviour is of larger risk. Outside of cultural differences, the same risks will also vary by industry, company, and AI application [2]. For example, the risk of unfair treatment is considerably more consequential for an AI system designed to assess loan applications than a recommendation system used within an e-commerce website.

AI risks are dynamic

Given AI's rapid pace of development and increasing applications, AI risks are constantly changing and evolving [3]. Therefore, identifying and monitoring AI risks needs to be an ongoing effort and any AI management framework will need to remain responsive to new and evolving risks [2], [5]. Additionally, various stages of the AI life cycle give rise to distinct challenges and risks. For example, the CEN-CENELEC Focus Group Report highlights new safety challenges emerging in the usage phase of AI systems that continue learning after their implementation is complete [20]. The report acknowledges the need for a methodology and tools that enable "verification, supervision, traceability, and recurring conformity assessment" of such systems as their functionality is modified after they are brought to the market [20, p. 11].

Key findings

- Transparency, privacy, security, safety, bias and accountability were the most common sources of risk identified within the literature surveyed. However, significant inconsistencies were found in how these risk factors were defined, conceptualised, and assessed. While guidelines and mitigation strategies have been proposed, evidence on their practical value and effectiveness is missing.
- In the absence of comprehensive risk management frameworks and structured processes for defining, assessing, and mitigating AI risks, an 'accountability gap' is emerging.
- Measurement of AI risks is particularly challenging due to unclear definitions, the inscrutable nature of AI systems and a complex ecosystem of third-party integrations.
- The rapid pace of AI development, as well as its dynamic nature, further complicates AI risk management efforts and exposes the need for adaptive frameworks that enable continuous monitoring of AI system safety throughout its entire lifecycle.
- Lastly, AI risks are context dependent and cultural differences are starting to emerge around proportionality of control measures. Many specific AI system risks are unique to their application context, which requires a different and yet coordinated approach in addressing them.

AI regulation, policy and strategy

In exploring AI regulation, policy and strategy, this literature review aims to map the current trends in government initiatives for AI management. To help extract the main challenges, our focus zooms in on four countries: the UK, the USA, the EU, and China.

The research questions were:

- What are the emerging trends in AI regulation?
- How are countries (UK, USA, EU, and China) approaching AI regulation?

The current policy discussion around AI is focused mainly on AI governance frameworks. Various publications in the field discuss differing regulatory and governance approaches to mitigating AI risks and harms and fostering innovation.

What are the emerging trends in AI regulation?

The current race in AI regulation

AI regulation is beginning to proliferate in different countries globally. The cost of non-intervention is a concern for governments that wish to make the most of the AI global market. Academics refer to AI regulation as a 'new playground' [22] for governments to support AI markets while addressing risks. Both academia and industry agree that AI regulations and policies are also a competitive tool at the geopolitical level due to their salient role in the political agenda [23], [24]. If one country or region can find a regulatory approach that is advantageous for companies operating within their territory while minimising disadvantages of AI systems for society, then these countries can position themselves as leading the AI race [22]. Within this context, both academic and policy papers refer to several countries already proposing AI governance measures to mitigate harms and provide incentives for companies developing AI applications. Nevertheless, most of these efforts are done in isolation by one country or entity, and due to factors such as cultural differences, the current regulatory landscape is fragmented.

AI risk regulation and the ethics gap

The literature and current governance frameworks propose that the starting point for AI regulation is in addressing risks and harms [22], [25], and governments endeavour to establish frameworks that protect citizens from AI-related harms. Although innovation is highlighted as a key aspect in the aforementioned AI race, the literature and governance activities indicate a trend toward risk-based regulation. Ethical use of AI dominates countries' AI governance approaches and is pursued by the EU, UK and the USA, to varying degrees.

While AI risk regulations aim to address ethics as well, academics have identified a gap between ethical guidelines and proposed legislation [26]. The manner in which AI regulation focuses on ethics does not usually match actual needs. This is because the ethical frameworks currently produced are difficult to put into practice [25]. One contributing factor is that laws are targeted to control human behaviour, and AI systems cannot be expected to comply with these laws. Instead, scholars have pushed for more human-centred AI (or 'human-in-the-loop') and to embed ethical codes or kill switches within the AI code.

The global landscape

The UK, EU, USA and China were chosen for the purposes of this literature review because they are deemed the current global leaders in AI. These countries and region have their own agendas, but fostering innovation is a common goal for each. This is because AI regulation can be a tool for market dominance. If a country's AI regulation is accepted to be the gold standard for AI regulation, entities operating within that country or region will benefit over those who are outside of it [24].

The United Kingdom

To date, the UK has produced key policy papers which shed light on how they would approach AI regulation.

In 2021, the UK government published the National AI Strategy [27] which lays out three key aims. They are to

1) invest and plan for the long-term needs of the AI ecosystem for UK to maintain its role as a leader in science and AI;
2) support the transition to an AI-enabled economy; and
3) ensure that the UK has appropriate governance at both the national and international level to foster innovation and investment and to protect the public and fundamental rights.

In 2022, the UK published the policy paper "Establishing a pro-innovation approach to regulating AI" [28] to operationalise the 2021 Strategy. This proposal seeks to provide clarity to businesses on how they can use and develop AI systems and, in turn, build consumer confidence regarding the safety and reliability of the systems [28]. However, this policy paper indicates that the UK is moving away from the EU approach stating that the proposed EU AI Act is not flexible enough, too centralised and limits innovation [28].

The European Union

The EU has been a leader in regulating the digital sector. It has recently proposed the EU AI Act which is thought to become the new GDPR for AI [29]. The proposed EU AI Act takes a horizontal

approach and encourages harmonised rules for the development, use and market placement of AI systems [30]. These AI systems are also subject to a predefined risk classification. There are four risk categories: unacceptable risk, high-risk, limited risks and minimal risks [31]. A conformity assessment system which is adapted from EU product safety law is also a requirement under this AI Act. There are two types of conformity tests, Internal Assessment and Notified Body Assessment, that need to be employed depending on the high-risk AI system in question [32]. Alongside this proposed EU AI Act, the EU has set out the New Legislative Framework (NLF) which endeavours to achieve a partnership between standards and regulation [33].

The proposed EU AI Act, however, faces heavy criticism. For one, the Members of the European Parliament (MEPs) argue that a balance must be maintained between security, privacy and innovation to avoid over-regulating the AI market as this may stifle innovation. Regardless, there are conflicting concerns about the risk to fundamental rights if there is no adequate regulation. Another crucial concern regards the scope of the proposed EU AI Act which covers both users and providers of AI systems [34]. The EU seems to have intentionally kept the language broad to be supplemented by standards, presumably produced under the NLF scheme. However, there are concerns that there may not be adequate reflection of minority views, such as SMEs, because such entities do not tend to take part in the technical committees that produce these standards [35].

The United States of America

The US government has set out a benchmark as to what trustworthy AI is [4]. The USA's National AI R&D Strategic Plan 2019 outlines eight strategic priorities, which include making long-term investments in AI research (Strategy 1); understanding and addressing the ethical, legal, and societal implication of AI (Strategy 3); and measuring and evaluating AI technologies through standards and benchmarks (Strategy 6), among others [16].

However, the USA's chosen approach is that specific regulatory guidelines are proposed on a agency-by-agency basis [36]. For example, the Food and Drug Administration (FDA) has released the Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan [37], and the Federal Trade Commission (FTC) has published a memo, Aiming for Truth, Fairness, and Equity in your company's use of AI [38], which lays out a roadmap for compliance expectations. The Department of Commerce has established a committee (the National Artificial Intelligence Advisory Committee (NAIAC)) to offer recommendations to the US President on how to approach the use and development of AI [39].

People's Republic of China

To maintain its position as a global leader in attracting AI investment, China has produced a Plan for Development of New Generation AI Technologies (AIDP) in 2017 to foster innovation [40]. It takes

a horizontal approach to China's AI capabilities as well as the risks and opportunities associated with them [4].

China has also begun discussions on privacy which indicates an emerging regulatory framework [41]. Some forecast that this framework will be stringent on classifying what constitutes sensitive personal information, even more so than the EU GDPR requirements [41]. The strict treatment of privacy in China may affect the long-term advancement of AI because the technology fundamentally revolves around data, some of which may be subject to privacy laws.

In other areas, China is seeking to regulate algorithmic systems. The Cyberspace Administration of China has passed draft regulations on design requirements standards and input data of algorithmic systems [42]. As China tends to have a huge impact in the AI field, its regulatory activities should be taken into account.

Key findings

- The number of countries working towards establishing their own AI governance approach has been increasing recently as per the salient role of AI in political agendas. Some are proposing regulation that address certain AI-related challenges and others are presenting national strategies and policies as a pathway for AI development.
- Governance efforts are often done in isolation, resulting in a fragmented approach requiring global alignment.
- There are numerous ethical frameworks addressing AI risks but they are difficult to translate into practice.

AI innovation

This literature review aims to understand drivers and barriers to novel AI development and adoption. It examines the mechanisms which spur AI innovation and those which lead to stagnation.

The research questions were:

- What are the drivers of AI innovation?
- What are the barriers to AI innovation?

Defining and contextualising AI innovation

Innovation in the field of AI, such as improvements in machine learning (ML), has the potential to improve lives and address complex global problems. While public discourse surrounding AI governance often focuses on minimising risks, many stakeholders also seek to drive responsible AI innovation, or at least remove barriers slowing the emergence of cutting-edge techniques and applications.

Innovation and adoption

Measurements of innovation are often blunt or incomplete. Instruments such as the number of relevant patents awarded [43] or the count of academic publications emerging over time [44] are not calibrated to examine the spread of innovative ML applications in business services or consumer products. Similarly, measures of AI adoption, like those collected per-sector or per-country [45], measure the distribution of technology while ignoring novelty. Therefore, to understand both drivers of and barriers to AI innovation, we embrace a definition of innovation which encompasses novelty and adoption. Limited evidence exists to assert a specific relationship between AI innovation and adoption. However, we hypothesise that novel AI techniques will improve adoption rates by solving common problems, and that increased adoption opens the door for innovation through sheer statistical probability.

We choose to avoid repeating common analyses showcasing global trends or predictions of innovation in AI (including overviews of emerging techniques and applications). Instead, we are focusing on the specific mechanisms of AI innovation. Barriers to innovation are explored, categorised by internal, external, and technical impediments. Standards provide one avenue to address barriers to AI innovation, though not the only path. Understanding the mechanisms of innovation and stagnation will inform what drivers and barriers fall within the scope of standards and which should be addressed by other methods of AI governance or market forces.

Mechanisms driving AI innovation

Models

The Technology Acceptance Model (TAM) is one of the most widely used empirical models of technological innovation and adoption. Focusing on the end user, TAM asserts that adoption is impacted by perceived usefulness, perceived ease of use, and attitude towards use [46].

The Diffusion of Innovations theory (DOI) expands its focus to include organisational mechanisms. These are time (including that required for implementation), the presence of channels of communication, cultural and social realities, and novelty or relative advantage [47].

The most comprehensive model examined, the Unified Theory of Acceptance and Use of Technology (UTAUT) is comprised of eight other models (including TAM and DOI) [48]. The UTAUT discusses expectation of effort, expectation of performance, social influence, and the removal of barriers impeding adoption [48].

AI-specific models of innovation and adoption are few and far between in the literature. However, we identified one model combining elements of novel technology generation and adoption. Fulton et al. build upon common models to introduce an AI-specific framework which incorporates temporal analysis of triggers of novel innovation from the past, present, and future [48]. The authors posit that the adoption of innovative AI technologies in industry is influenced by market forces like customer needs and relative advantage, as well as external and internal factors like a changing environment and available resources [48].

The common use of the aforementioned models [48] indicates innovation and adoption occur at the intersection of individual, organisational, and systemic factors. Mechanisms of innovation often play out at multiple levels. Each model makes note of the social nature of innovation, whether through individual attitude or broader social influence.

Despite this, there is a gap in the literature regarding the specific role that social and cultural factors play in AI innovation.

Organisational need

Organisational need drives improvement and innovation of AI technologies. The benefits gained from improvement in these areas facilitate a large market of potential customers seeking AI solutions which can offer said benefits. The presence of such a market drives novel AI solutions, improvements and adoption.

First, factors which provide a competitive advantage are key. These include reduced time to insight (speed) [48]; savings from the efficiency of technologies like chatbots or screen assist (cost) [48]; use of personalisation to drive revenue and expand engagement (customisation) [49]; and operational efficiency and management through AIOps (scalability/reusability) [49], [50].

Second, organisations seek to minimise liability and risk. Even if a solution provides a competitive advantage or meets a need, the solution is not viable if it introduces relatively greater risk or harm. Triggers of innovation in this space are numerous, but include a need to prevent harm, especially for marginalised groups (inclusivity) [48]; a need for solutions to integrate and understand disparate systems to prevent unintended consequences (interdisciplinarity) [48], [51]; and, more broadly, market demand for ethical AI development and deployment strategies (responsibility) [49].

Technological breakthroughs and limitations

Innovation in AI techniques creates paths for aggregating innovation through completely new AI applications, or through applications improved enough for integration into consumer products or services. For example, the evolution of deep learning techniques has led to newly viable applications in computer vision, natural language processing and generation, classification, and perception [52]. Technological advancement in non-AI fields can fuel AI innovation as well. Experts surveyed by IBM [52] say quantum computing will have an outsized impact on AI innovation.

In addition to breakthroughs, technical restrictions can drive new techniques and applications when developers are forced to solve problems in a new way. The limited availability of certain data has led to more data-efficient techniques like one-shot learning [52]. The data landscape which so hugely influences AI innovation has been turned on its head, with new applications emerging for small datasets [49], [52].

In this way, innovation can beget further innovation. Promoting AI innovation could lead to compounding returns, although this is not a guarantee as plateaus do occur.

Barriers to AI innovation

Internal barriers

Literature [53], [54], on the mechanisms of innovation highlights three barrier types, or 'lock-ins', as impeding AI innovation - technological, behavioural, and institutional lock-ins. During the AI development phase and beyond, gathering and structuring data is fundamental [55], which means that infrastructure suitable for AI technology is required [56]. Existing IT infrastructure is not necessarily appropriate for new technology due to incompatibility between systems, but users tend to stick with incumbent technology. Their tendency impedes the introduction of new technologies and, thus, innovation [53]. This is the so-called 'technological lock-in', and evidence [57] suggests it impacts innovation.

In considering the human factor, users of incumbent technology are also sometimes reluctant to adjust their behaviour to new technologies. This is defined as the 'behavioural lock-in' [54]. Behavioural lock-in could hinder the development of new infrastructure. In order to promote AI innovation, it is also important to change the mindset of employees along the AI lifecycle.

Lastly, radical transformation such as AI adoption entails set-up or fixed-costs [53], and it is crucial for companies to have sufficient financial resources and human capital for successful AI adoption. The lack of these resources makes it difficult to transform the current system infrastructure and deploy new technology. This is the so-called 'institutional lock-in' [53], and SMEs tend to face this

barrier due to their insufficient capacity [56], [57]. A survey [57] conducted in the UK demonstrates the significant difference in AI adoption between SMEs and large companies, and institutional lock-in might be one culprit.

AI innovation is affected by not only technical capability but also individual and organisational capacity. Even if a company is motivated to improve business practices or products through deploying and developing AI technology, these lock-ins could prevent the company from achieving this goal. Despite significant challenges on SMEs caused by these lock-ins, most studies in support of these companies have focused on the necessity of budget.

External barriers

As data is the foundation of AI technology, the quantity and quality of data affects the success of AI applications [58]. Data exchange across sectors and nations could improve the accuracy of such systems, but sector-based regulations [59] and countries' privacy laws prevent AI developers from accessing and gathering data [56]. This reality introduces major barriers to innovation.

Furthermore, the contrast between national AI strategies could also hinder global trade of AI technology [1]. For example, while the US and UK tend to prioritise AI innovation, the EU and China emphasise the mitigation of risks caused by the technology. This implies that new AI technology developed in British or American companies may not be allowed in European or Chinese companies due to the concerns of potential risks. The delay of AI adoption could be an obstacle to AI innovation.

Key findings

- Innovation is impacted by individual, organisational, social/cultural, and systemic factors. More research is needed to understand the mechanisms of AI innovation, particularly social and cultural factors.

- Organisational needs, like a desire to remain competitive, respond to market demand, and minimise risk and liability, foster AI innovation.

- Technological breakthroughs can produce further innovation, and some technological restrictions spur innovation as practitioners seek new ways to solve problems. However, other technological factors like legacy IT infrastructure impede innovation.

- Human behaviour can act as a barrier to innovation, as integration of AI requires training and education which may be missing on a large scale. The human education pipeline and a lack of talent, particularly cross-functional talent, limit AI innovation.

- Inconsistency between national strategies and technologies themselves prevent innovation through a lack of interoperability or mismatched compliance requirements.

The role of AI standards

Introduction

Research on AI standards is still limited. While AI standardisation efforts started several years ago, there are few academic papers [1], [60] and reports [61] that describe AI standardisation efforts.

Most publications refer to the work of international standards bodies. The work of ISO and IEEE is mentioned in a majority of articles identified [1], [60]. These bodies are described as playing a key role in the wider AI governance ecosystem [1]. One of their strengths is their institutional capacity to create global standards by achieving consensus. International SDOs also are a reference point for other stakeholders, including national standards bodies, governments and industry.

Most research in the field is related to the role of standards in emerging technologies. A recent publication from the UK government is one example [62]. Furthermore, current literature illustrates that ethics plays a pivotal role in standardisation work around emerging technologies [63]. Documents consulted point to the capacity of AI standards to build trust and address risks in the initial phase of AI system design [63].

Despite limited literature analysing AI standards, our research process allowed us to uncover several roles standards are likely to play in the AI governance sphere. Standards are particularly well suited for managing interoperability, providing procedural guidance to organisations, and assuring the quality and safety of AI products and services. Standards are also crucial for supporting regulation and enabling compliance with legal obligations.

AI standards for mitigating risks and harms

Standards bodies have studied the impact of AI systems and proposed a number of solutions to resulting challenges, particularly on potential risks and harm they impose on individuals, organisations, society and existing systems [62]. Consultancy firms, such as Ernst & Young (EY), note how these risks have increasingly been addressed by many countries and organisations through risk-based approaches to AI governance [4].

Risks and harms are concerning, especially due to AI systems' rapid development and uptake. Currently, such systems are commonly utilised to resolve highly complex tasks such as autonomous driving, image and speech recognition [10]. Many AI systems for autonomous vehicles are in development by major companies such as Tesla, Inc. However, the ever-increasing number of accidents caused by these AI systems [64] [66] have prompted both members of the public and regulators to demand more from system developers. At this junction, regulators are weighing a soft or hard law approach, the latter concerning some for its potentially adverse consequences on innovation and economic prosperity.

Against this backdrop, standards are often presented as appropriate tool to manage AI risks and harm because they can be voluntary and require broad agreement on common principles among differing national bodies. As noted by an EY report, there is presently a valuable opportunity for alignment between global regulators on risk management strategies given the general agreement on responsible AI principles [4].

AI standards to complement regulation

Literature on AI regulation and policy discusses the complexity of defining AI for regulatory purposes, and asserts that the breadth of AI applications make it extremely difficult to approach governance with a single regulatory framework or policy initiative [20]. Using multiple policy instruments is considered the way forward for better AI governance [67]. In this scenario, standards are often presented as governance mechanisms that may support AI regulation and the challenges encountered [67]. Some scholars observe that in the field of AI governance, attention is increasingly shifting from governments to standards-making bodies [23].

Literature concerning standards stressed their function in assisting government initiatives [68]. They can play a key role in operationalising some of the principles and processes embedded in policy and regulation.

Nevertheless, the literature also shows that standards, while they can be technical in nature, have implications for governments and organisations [23]. For example, early-stage standardisation in AI could help inform regulatory approaches and reach a consensus within industry [62]. Standardisation efforts can also be a form of enterprise self-regulation and lead to reducing the need for stringent government regulation [62].

At a national level, several governments are taking the lead in the race for AI regulation [20]. Besides proposed policy changes to encourage AI development, these documents are increasingly prioritising engagement in standardisation processes and collaboration with international standard-setting bodies [20]. In this regard, standards could support global solutions required due to the transboundary nature of AI systems.

AI standards in innovation

AI standards promote innovation by enabling increased compatibility among application programming interfaces (APIs) [55]. As compatibility and interconnectivity could also decrease the transparency of AI systems, scholars [48], [69], mentioned that standards could help maintain transparency and trustworthiness. This could be done by encouraging the process for evaluation and guidelines on requirements for explainability and reliability [70].

Standards generally, and management standards in particular, can also address institutional barriers to AI innovation through changing companies' activities. For example, previous research has shown that the adoption of management standards leads to changes in employee mindset and an increase in company revenue [71]. One academic [1], highlights that standards could become another tool for entrepreneurs that wish to promote a culture of transparency and safety in AI. Still, further research reflecting on how well management standards perform would be valuable.

However, if not done with care, standardisation could obstruct innovation due to inflexible or quickly outdated standards [62]. The literature presents a potential conflict of interest amongst stakeholders in regard to innovation. Start-ups are focused on leveraging AI technologies to launch new products and services and could be negatively impacted by mandatory standards. Large and well-resourced companies are more open to standards that could reduce uncertainty, provide operational guidelines, and mitigate risks for their consumers [62].

Key findings

Existing literature on AI standards is scarce. Some publications address the role of standards in relation to emerging technologies that could be applied to AI. Concretely, most publications found describe the work of international standardisation [1], [60]. Nevertheless, in relation to our three research clusters, the following roles were identified:

- Standards can help manage AI risks and harms by unifying terminologies, clarifying concepts, and operationalising principles like transparency, fairness, and reliability.

- In light of the emerging government initiatives, standards have received increased attention as a method of increasing global alignment of AI. Though they also are fundamental in supporting regulation, a gap exists between standards efforts and policy objectives [1].

- Standards can help promote innovation through improving compatibility among systems and guiding organisational processes and procedures, including for pursuing transparency and trustworthiness. However, the impact depends upon the capacity of companies to apply standards to their operations.

Conclusion

Given the horizontal nature of AI and its rapid evolution, governance mechanisms are crucial to protect individuals, organisations and society. Through this literature review, we aim to obtain an understanding of the requirements of AI governance by focusing on AI risks and harms, AI innovation, and AI regulation, policy and strategy. We also hope to explore how standards can play a role in the challenges revealed by each research cluster in the coming years. The main challenges and issues uncovered are summarised below.

AI risk and harms: Although AI brings many opportunities, it also gives rise to a host of unintended consequences that threaten fundamental rights and public safety. This review has showcased that transparency, privacy, security, safety, bias and accountability were the most common sources of risk discussed within the literature. However, terminologies and definitions around these risks were often incohesive and fragmented between publishing fields. Additionally, literature points to a lack of comprehensive risk management frameworks and structured processes for assessing AI risks, leading to unclear accountability structures throughout the AI value chain. Measurement of AI risks was found to be particularly challenging due to the novelty and complexity of AI systems, as well as missing measurement techniques and processes.

AI regulation, policy and strategy: This literature review has showcased that countries worldwide are developing their own AI governance policies, strategies and, in some cases, regulations.

However, countries' approaches differ: some prioritise innovation while others place focus on risk approaches, or a combination of both. This has resulted in a fragmented AI governance landscape. Four countries leading the AI regulation 'race' were further examined to provide insight into emerging regulation to inform our research. Finally, many of the ethical frameworks that have been published thus far have been difficult to operationalise.

AI innovation: Throughout this review we have identified drivers of and barriers to AI innovation, including AI adoption. One of the key drivers to AI innovation is the organisational need to maintain competitiveness in the market by mitigating AI risk and harms and reducing resulting product failures. Technological breakthroughs can also promote innovation through aggregated benefits. Models of technological innovation and uptake suggest both are affected by individual, organisation, social/cultural and systemic factors. Internal barriers to AI innovation consist of technological, institutional and behavioural lock-ins, whereas external barriers arise from inconsistency between national AI strategies and technologies which impede AI innovation due to a lack of interoperability or unclear or contradictory compliance requirements.

The role of AI standards: While we found limited literature on AI standards, this review also clarified the role standards play in AI governance. For example, standards unify terminologies and clarify concepts around AI risk. Standards can help organisations manage risk and meet regulatory requirements, but better-resourced organisations access standards more easily. The fragmented global AI governance landscape, with several policies and regulations emerging globally, could be addressed in part by standards. They can play a role in operationalising transparency and trustworthiness principles, one of the many challenges organisations face.

References

[1] P. Cihon, 'Technical Report, Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development', *Future Humanity Inst. Univ. Oxf.*, 2019.
[2] B. Cheatham, K. Javanmardian, and H. Samandari, 'Confronting the risks of artificial intelligence', *McKinsey Q.*, vol. 2019, no. 2, 2019, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079283855&partnerID=40&md5=86169ee128e7c7f33dbf9659a97fda36
[3] B. W. Wirtz, J. C. Weyerer, and I. Kehl, 'Governance of artificial intelligence: A risk and guideline-based integrative framework', *Gov. Inf. Q.*, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126090396&doi=10.1016%2fj.giq.2022.101685&partnerID=40&md5=ba7bab1e031c65b0fd444ce43f61599f
[4] A. Koene, G. Ezeani, R. Kumar, N. Santiago, and D. Wright, 'A survey of artificial intelligence risk assessment methodologies. The global state of play and leading practices identified', Trilateral Research and Ernst & Young (EY), 2022. [Online]. Available: https://www.trilateralresearch.com/wp-content/uploads/2022/01/A-survey-of-AI-Risk-Assessment-Methodologies-full-report.pdf
[5] National Institute of Standards and Technology, 'AI Risk Management Framework: Initial Draft'. NIST, 2022. [Online]. Available: https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf
[6] R. Rodrigues, 'Legal and human rights issues of AI: Gaps, challenges and vulnerabilities', *J. Responsible Technol.*, vol. 4, p. 100005, Dec. 2020, doi: 10.1016/j.jrt.2020.100005.
[7] ART/1 Artificial Intelligence, 'Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations', The British Standards Institution (BSI), International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), BS ISO/IEC 38507:2022, May 2022. [Online]. Available: https://bsol.bsigroup.com/Bibliographic/BibliographicInfoData/000000000030379288
[8] K. Jia and N. Zhang, 'Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines', *Electron. Mark.*, vol. 32, no. 1, pp. 59–71, 2022.
[9] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), 'Draft BS ISO/IEC 23894 Information technology - Artificial intelligence - Risk management.' International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), Jan. 21, 2022.
[10] A. Steimers and M. Schneider, 'Sources of Risk of AI Systems', *Int. J. Environ. Res. Public. Health*, vol. 19, no. 6, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126452843&doi=10.3390%2fijerph19063641&partnerID=40&md5=3c009655eaebb1c6f2fe1c7136b906e6
[11] S. Wachter, B. Mittelstadt, and L. Floridi, 'Transparent, explainable, and accountable

AI for robotics', *Sci. Robot.*, vol. 2, no. 6, 2017, doi: 10.1126/scirobotics.aan6080.
[12] C. Bartneck, C. Lütge, A. Wagner, and S. Welsh, *Risks in the Business of AI*. ["HIT Lab NZ, University of Canterbury, Christchurch, New Zealand", 'Institute for Ethics in Artificial Intelligence, Technical University of Munich, München, Germany', 'College of Engineering, Pennsylvania State University, University Park, PA, United States', 'Department of Philosophy, University of Canterbury, Christchurch, New Zealand']: Springer Nature, 2021. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103708666&doi=10.1007%2f978-3-030-51110-4_6&partnerID=40&md5=a8871cbaecd26cf376d2308ea98996f2
[13] F. Z. Borgesius, 'Discrimination, artificial intelligence, and algorith-mic decision-making', Council of Europe, 2018. [Online]. Available: https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73
[14] J. Krijger, 'Enter the metrics: critical theory and organizational operationalization of AI ethics', *AI Soc.*, 2021, doi: 10.1007/s00146-021-01256-3.
[15] European Commission, 'Ethics guidelines for trustworthy AI', European Commission, Apr. 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[16] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi, 'Operationalising AI ethics: barriers, enablers and next steps', *AI Soc.*, 2021, doi: 10.1007/s00146-021-01308-8.
[17] National Institute of Standards and Technology (NIST), 'AI Risk Management Framework Concept Paper'. National Institute of Standards and Technology (NIST), Dec. 13, 2021. [Online]. Available: https://www.nist.gov/system/files/documents/2021/12/14/AI%20RMF%20Concept%20Paper_13Dec2021_posted.pdf
[18] A. Steimers and M. Schneider, 'Sources of Risk of AI Systems', *Int. J. Environ. Res. Public. Health*, vol. 19, no. 6, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126452843&doi=10.3390%2fijerph19063641&partnerID=40&md5=3c009655eaebb1c6f2fe1c7136b906e6
[19] Lilian Edwards, 'Regulating AI in Europe: four problems and four solutions', Ada Lovelace Institute, Expert Opinion, Mar. 2022. [Online]. Available: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf
[20] CEN CENELEC, 'CEN-CENELEC Focus Group Report - Road Map on Artificial Intelligence (AI)', CEN CENELEC, 2020. [Online]. Available: https://www.standict.eu/node/4854
[21] A. Luusua and J. Ylipulli, 'Artificial intelligence and risk in design', in *DIS Companion - Companion Publ. ACM Des. Interact. Syst. Conf.*, ["University of Oulu, Oulu, Finland", 'Aalto University, Espoo, Finland'], 2020, pp. 1235–1244. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090503450&doi=10.1145%2f3357236.3395491&partnerID=40&md5=e2165f582ad1f38

921d57648ccd4da42
[22] N. A. Smuha, 'From a "race to AI" to a "race to AI regulation": regulatory competition for artificial intelligence', *Law Innov. Technol.*, vol. 13, no. 1, pp. 57–84, 2021.
[23] L. Schmitt, 'Mapping global AI governance: a nascent regime in a fragmented landscape', *AI Ethics*, vol. 2, 2021, doi: 10.1007/s43681-021-00083-y.
[24] Duncan Stewart, Paul Lee, Ariane Bucaille, and Gillian Crossan, 'AIs wide shut: AI regulation gets (even more) serious, AI in 2022 will face intensifying regulatory scrutiny, with implications that resonate across industries'. Deloitte, Dec. 01, 2021. [Online]. Available: https://www2.deloitte.com/xe/en/insights/industry/technology/technology-media-and-telecom-predictions/2022/ai-regulation-trends.html
[25] D. Gervais, 'Towards an effective transnational regulation of AI', *AI Soc.*, 2021, doi: 10.1007/s00146-021-01310-0.
[26] P. Almeida, C. Santos Jr, and J. Farias, 'Artificial Intelligence Regulation: a framework for governance', *Ethics Inf. Technol.*, vol. 23, Sep. 2021, doi: 10.1007/s10676-021-09593-z.
[27] Secretary of State for Digital, Culture, Media and Sport, 'National AI Strategy'. Secretary of State for Digital, Culture, Media and Sport, Sep. 22, 2021. [Online]. Available: https://www.gov.uk/government/publications/national-ai-strategy
[28] Rt Hon Nadine Dorries MP, 'Policy paper Establishing a pro-innovation approach to regulating AI', Digital, Culture, Media and Sport (DCMS), London, Policy Paper, Jul. 2022. [Online]. Available: https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement
[29] Jane Finlayson Brown, 'Key provisions of the Draft AI Regulation'. Allen & Overy, May 24, 2021. [Online]. Available: https://www.allenovery.com/en-gb/global/news-and-insights/publications/key-provisions-of-the-draft-ai-regulation
[30] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act', 2021, doi: 10.9785/cri-2021-220402.
[31] European Commission, 'Regulatory framework proposal on artificial intelligence'. European Commission, Jun. 07, 2022. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[32] Dechert LLP, 'European Commission's Proposed Regulation on Artificial Intelligence: Conducting a Conformity Assessment for High-Risk AI - Say What?' Dechert LLP, Nov. 16, 2021. [Online]. Available: https://www.dechert.com/knowledge/onpoint/2021/11/european-commission-s-proposed-regulation-on-artificial-intellig.html
[33] Mark McFadden, Kate Jones, Emily Taylor, and Georgia Osborn, 'Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation', Oxford Information Labs, Dec. 2021. [Online]. Available: https://oxil.uk/publications/2021-12-02-oxford-internet-institute-oxil-harmonising-ai/Harmonising-AI-OXIL.pdf
[34] Lilian Edwards, 'Expert explainer: The EU AI Act Proposal'. Ada Lovelace Institute,

Apr. 2022. [Online]. Available: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf
[35] Laurie Clark, 'MEPs are preparing to debate Europe's AI Act. These are the most contentious issues.' Tech Monitor, Sep. 28, 2021. [Online]. Available: https://techmonitor.ai/policy/meps-are-preparing-to-debate-europes-ai-act-these-are-the-most-contentious-issues
[36] Heather Sussman, 'U.S. Artificial Intelligence Regulation Takes Shape'. Orrick, Nov. 18, 2021. [Online]. Available: https://www.orrick.com/en/Insights/2021/11/US-Artificial-Intelligence-Regulation-Takes-Shape
[37] US Food and Drug Administration (FDA), 'Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan'. US Food and Drug Administration (FDA), Jan. 2021. [Online]. Available: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
[38] Elisa Jillson, 'Aiming for truth, fairness, and equity in your company's use of AI', *Business Guidance*, Apr. 19, 2021. https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai
[39] US Department of Commerce, 'Department of Commerce Establishes National Artificial Intelligence Advisory Committee'. US Department of Commerce, Sep. 08, 2021. [Online]. Available: https://www.commerce.gov/news/press-releases/2021/09/department-commerce-establishes-national-artificial-intelligence
[40] A. Saveliev and D. Zhurenkov, 'Artificial intelligence and social responsibility: the case of the artificial intelligence strategies in the United States, Russia, and China', *Kybernetes*, vol. 50, no. 3, pp. 656–675, Jan. 2021, doi: 10.1108/K-01-2020-0060.
[41] H. Roberts, J. Cowls, J. Morley, M. Taddeo, V. Wang, and L. Floridi, 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation', *AI Soc.*, vol. 36, no. 1, pp. 59–77, Mar. 2021, doi: 10.1007/s00146-020-00992-2.
[42] Ada Lovelace Institute, 'Regulate to innovate: A route to regulation that reflects the ambition of the UK AI Strategy'. Ada Lovelace Institute, Nov. 2021. [Online]. Available: https://www.adalovelaceinstitute.org/report/regulate-innovate/
[43] H. Habibollahi Najaf Abadi and M. Pecht, 'Artificial Intelligence Trends Based on the Patents Granted by the United States Patent and Trademark Office', *IEEE Access*, vol. 8, pp. 81633–81643, 2020, doi: 10.1109/ACCESS.2020.2988815.
[44] Daniel Zhang *et al.*, 'The AI Index 2022 Annual Report', AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, Mar. 2022. [Online]. Available: https://aiindex.stanford.edu/report/
[45] International Business Machines (IBM), 'IBM Global AI Adoption Index 2022', International Business Machines (IBM), 2022. [Online]. Available: https://www.ibm.com/watson/resources/ai-adoption
[46] H. Taherdoost, 'A review of technology acceptance and adoption models and theories', *11th Int. Conf. Interdiscip. Eng. INTER-ENG 2017 5-6 Oct. 2017 Tirgu Mures Rom.*, vol.

22, pp. 960–967, Jan. 2018, doi: 10.1016/j.promfg.2018.03.137.
[47] I. Sila, 'The state of empirical research on the adoption and diffusion of business-to-business e-commerce', *Int. J. Electron. Bus.*, vol. 12, no. 3, pp. 258–301, Jan. 2015, doi: 10.1504/IJEB.2015.071386.
[48] R. Fulton, D. Fulton, and S. Kaplan, 'Artificial Intelligence: Framework of driving triggers to past, present and future applications and influencers of industry sector adoption'. arXiv, 2022. doi: 10.48550/ARXIV.2204.01518.
[49] M. Rimol, 'Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation', Gartner, Sep. 2021. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2021-09-07-gartner-identifies-four-trends-driving-near-term-artificial-intelligence-innovation
[50] T. Schmid, W. Hildesheim, T. Holoyad, and K. Schumacher, 'The AI Methods, Capabilities and Criticality Grid', *KI - Künstl. Intell.*, vol. 35, no. 3, pp. 425–440, Nov. 2021, doi: 10.1007/s13218-021-00736-4.
[51] Y. K. Dwivedi *et al.*, 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy', *Int. J. Inf. Manag.*, vol. 57, p. 101994, Apr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.
[52] International Business Machines (IBM), 'The new AI innovation equation'. International Business Machines (IBM). [Online]. Available: https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-innovation-equation.html#section1
[53] T. J. Foxon, 'Technological and institutional " lock-in " as a barrier to sustainable innovation', 2002. [Online]. Available: https://www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/icept/7294726.PDF
[54] G. Cecere, N. Corrocher, C. Gossart, and M. Ozman, 'Lock-in and path dependence: an evolutionary approach to eco-innovations', *J. Evol. Econ.*, vol. 24, no. 5, pp. 1037–1065, Nov. 2014, doi: 10.1007/s00191-014-0381-5.
[55] U. Paschen, C. Pitt, and J. Kietzmann, 'Artificial intelligence: Building blocks and an innovation typology', *Artif. Intell. Mach. Learn.*, vol. 63, no. 2, pp. 147–155, Mar. 2020, doi: 10.1016/j.bushor.2019.10.004.
[56] Andrew Evans and Anja Heimann, 'AI ACTIVITY IN UK BUSINESSES: An assessment of the scale of AI activity in UK businesses and scenarios for growth over the next twenty years', Department for Digital, Culture, Media, and Sport (DCMS), Jan. 2022. [Online]. Available: https://www.gov.uk/government/publications/ai-activity-in-uk-businesses
[57] Department for Digital, Culture, Media and Sport (DCMS), 'Data foundations and AI adoption in the UK private and third sectors: Executive Summary'. Department for Digital, Culture, Media and Sport (DCMS), Aug. 16, 2021. [Online]. Available: https://www.gov.uk/government/publications/data-foundations-and-ai-adoption-in-the-uk-private-and-third-sectors/data-foundations-and-ai-adoption-in-the-uk-private-and-third-sectors-executive-summary
[58] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, 'Applications of Artificial

Intelligence in Transport: An Overview', *Sustainability*, vol. 11, no. 1, 2019, doi: 10.3390/su11010189.
[59] Department for Digital, Culture, Media and Sport (DCMS), 'Data: A new direction'. Department for Digital, Culture, Media and Sport (DCMS), Sep. 10, 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document__Accessible_.pdf
[60] T. Zielke, 'Is Artificial Intelligence Ready for Standardization?', in *Commun. Comput. Info. Sci.*, 2020, vol. 1251 CCIS, pp. 259–274. doi: 10.1007/978-3-030-56441-4_19.
[61] European Commission, Joint Research Centre, S. Nativi, and S. De Nigris, *AI Watch, AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework*. Publications Office, 2021. doi: 10.2760/376602.
[62] N. Brown *et al.*, 'The Role of Standardisation in Support of Emerging Technologies in the UK', Department for Business, Energy and Industrial Strategy (BEIS), London, May 2022. [Online]. Available: https://www.gov.uk/government/publications/role-of-standardisation-in-support-of-emerging-technologies
[63] I. Mezgár and J. Váncza, 'From ethics to standards – A path via responsible AI to cyber-physical production systems', *Annu. Rev. Control*, vol. 53, pp. 391–404, Jan. 2022, doi: 10.1016/j.arcontrol.2022.04.002.
[64] S. Gray, 'List of Driverless Vehicle Accidents', *ITGS News*, 2018. [Online]. Available: https://www.itgsnews.com/list-of-driverless-ve hicle-accidents/
[65] B. Pietsch, 'Killed in Driverless Tesla Car Crash, Officials Say', *New York Times*, 2021. [Online]. Available: https://www.nytimes.co m/2021/04/18/business/tesla-fatal-crash-texas.html
[66] D. Wakabayashi, 'Self-Driving Uber Car Kills Pedestrian in Arizona', *New York Times*, 2018. [Online]. Available: https://www.nytime s.com/2018/03/19/technology/uber-driverless-fatality.html
[67] R. Clarke, 'Regulatory alternatives for AI', *Comput. Law Secur. Rev.*, vol. 35, no. 4, pp. 398–409, Aug. 2019, doi: 10.1016/j.clsr.2019.04.008.
[68] Laura Degallaix and Mariolina Eliantonio, 'THE USE OF STANDARDS IN LEGISLATION AND POLICIES: How to ensure a legally-sound and socially-acceptable use of standards by policy-makers in view of the CJEU James Elliott ruling and other court cases', ECOS - European Environmental Citizens Organisation for Standardisation, Jun. 2018. [Online]. Available: https://ecostandard.org/wp-content/uploads/2018-06-11-The-use-of-standards-in-legislation-and-policies-ECOS-discussion-paper.pdf
[69] SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE, 'THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE', NATIONAL SCIENCE AND TECHNOLOGY COUNCIL (USA), Jun. 2019. [Online]. Available: https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf
[70] National Institute of Standards and Technology (NIST), 'U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools', National Institute of Standards and Technology (NIST), Aug. 2019. [Online]. Available:

https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf
[71] A. Peršič, M. Markič, and M. Peršič, 'The impact of socially responsible management standards on the business success of an organisation', *Total Qual. Manag. Bus. Excell.*, vol. 29, no. 1–2, pp. 225–237, Jan. 2018, doi: 10.1080/14783363.2016.1174059.

ANNEX C - Standards-mapping

Description

We conducted a mapping exercise to visualise AI standardisation efforts led by ISO and IEEE over the last four years (2018-2022). AI-related standards were divided into foundational standards (those that clarify concepts and form the basis of other standards specifications) and implementation standards (those which contain specific processes or requirements for developing or adopting AI).

The mapping of published AI-related standards was included in the report (*The AI Standardisation Landscape*). However, we also mapped AI standards in progress, seen below (***Figure 5***).

Map of draft AI standards

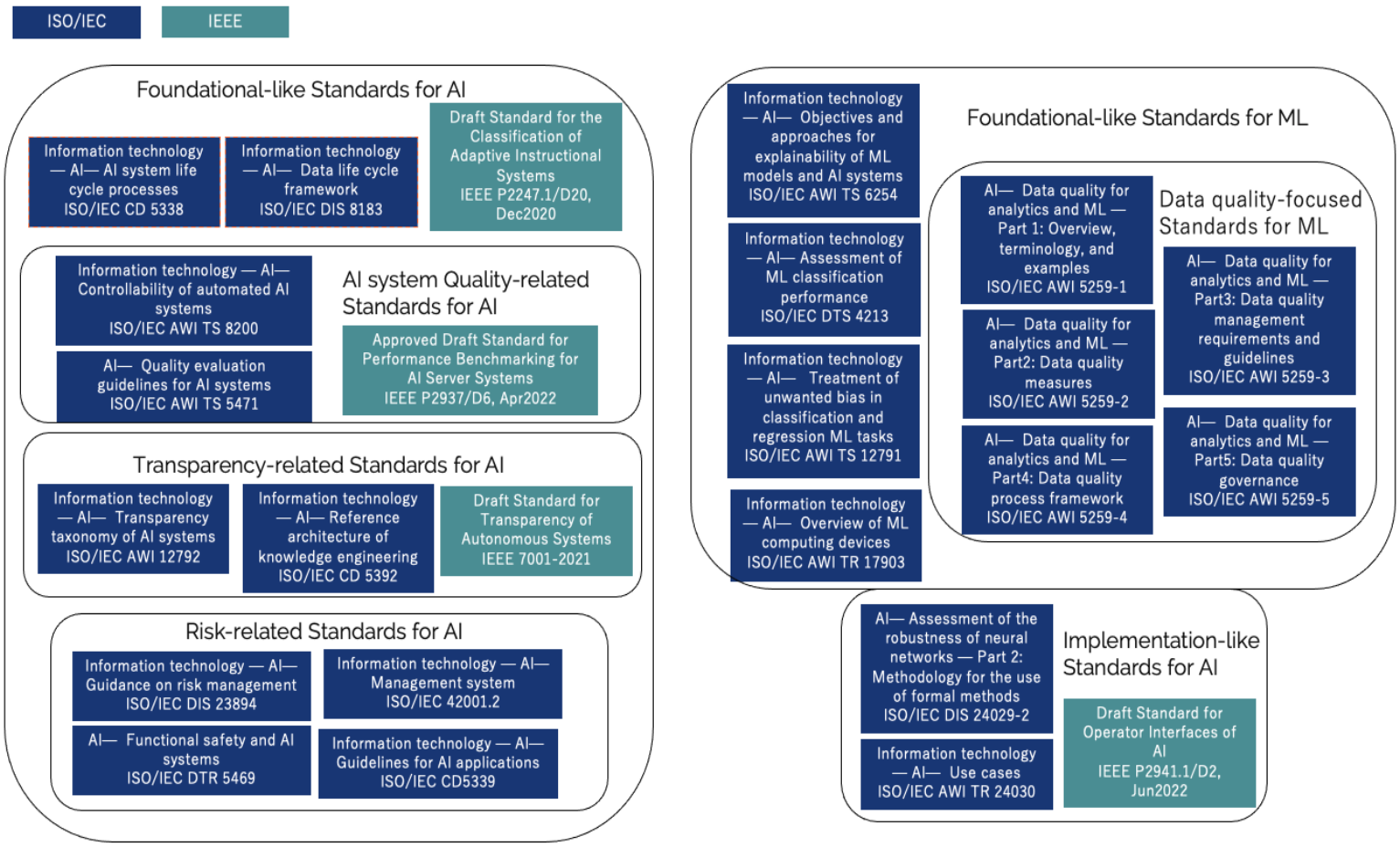


***Figure 5**: AI standards under development*

ANNEX D - Drivers of change

The governance of emerging technology is, by its nature, in a state of flux. Therefore, research on this subject should not only be future-facing but also adaptive and iterative. Within this space, identifying drivers of change can be a useful strategy [1].

In their 2030 strategy, ISO identified four primary drivers of change (Economic, Technology, Society, and Environment) with the goal of preparing their organisation for disruptive changes and adapting quickly to them [1]. While the four drivers are useful for strategising their future standards work, they are limited as they proactively define the categories drivers of change that may affect future standards before collecting events and triggers.

Our process of horizon scanning differed. First, we used appropriate methods to gather and generate drivers of change, and then subsequently categorised them by theme. In this way, we ensured the production of drivers of change were not subject to predefined criteria which could inadvertently bias their collection.

We have collected, analysed and plotted 47 drivers of change within five categories. They were derived through multiple sources: the literature review of academic and non academic sources, semi-structured interviews and a workshop with experts and stakeholders. While plotted drivers cannot form a complete picture of all forces of change, or display interplay between them, they are a useful tool for guiding further research.

Reference

[1] International Organisation for Standardisation (ISO). "ISO Strategy 2030." International Organisation for Standardisation (ISO), 2021. https://www.iso.org/publication/PUB100364.html.

Drivers of change list

Category: Regulatory/Governmental

Driver	Timeline
• AI policies and strategies • Bias bounties • Multi-lateral governance • Global alignment • Sandboxes • Vertical sector restrictions • Responsible AI requirements	1 – 2 years
• AI adopters (SMEs) in standards process • AI regulation • Demand for assurance • Ethics auditing • Designated standards • Demand for compliance services	2 – 3 years
• Global regulatory market • Compliance and quality signal (kitemark) • Context-specific metrology • Regulatory requirements for user-wellbeing	3 – 5 years

• Regulatory requirements for user well-being	5 – 10 years

Category: Technological

Driver	Timeline
• Blockchain	1 – 2 years
• Federated architecture • Web 3.0 • Connectivity (5G, etc.) • Integration with non-AI systems • Hardware improvements	2 – 3 years
• Battery technologies • Metaverse • Privacy enhancing technologies	3 – 5 years
• Quantum computing • Artificial General Intelligence (AGI) • Symbolic and sub-symbolic hybridization	5 – 10 years

Category: Economic

Driver	Timeline
• Economic recession • Decrease in open-source models • Government incentives	1 – 2 years
• Innovative business models	2 – 3 years

Category: Data-driven

Driver	Timeline
• Data origin and re-identification	1 – 2 years
• Data curation • Data quality and governance • Small and wide data	2 – 3 years

Category: Socio-technical

Driver	Timeline
• Publicised AI safety incidents • Consumer demand for responsible AI • Talent scarcity • Cybersecurity	1 – 2 years
• Sector-specific implications of trustworthy AI • AI taxonomy adoption • Large-scale AI failures	2 – 3 years
• Social cost of automation • AI carbon footprint • Quantum ethics	3 – 5 years