Summer Internship Project Report

on

STUDY OF ARTIFICIAL INTELLIGENCE REGULATIONS TO IDENTIFY STANDARDIZATION OPPORTUNITIES

Author - Oishika Datta

ACKNOWLEDGEMENT

I express my heartfelt gratitude to The Bureau of Indian Standards for providing me with this invaluable opportunity to learn and grow. The support, guidance, and exposure to real-world challenges have been truly enriching and instrumental in shaping my professional development.

I would like to express my sincere gratitude to my mentor, Mr. Kshitij Bathla, Scientist C, for his exceptional mentorship and steadfast support throughout the entirety of this project. His profound insights into global developments in the field have not only broadened my perspective but also deepened my understanding of the complexities involved. His guidance has been pivotal in navigating challenges and shaping the course of this endeavor, making it a truly enriching learning experience. I am extremely grateful for his patience and encouragement.

I am grateful to Mrs Reena Garg, Scientist G and HoD - LITD, as well as the entire Electronics and IT Department at BIS, and my fellow interns, for their invaluable as well as the entire for their collaboration and support, and for fostering a conducive and collaborative environment.

The insights I gained from Mr. Gautam Banerjee and his expert team at Business Brio, have helped me enhance my understanding of the industry's operations. Their invaluable guidance and firsthand knowledge have been crucial in shaping this report.

I would also like to extend my heartfelt thanks to my esteemed professors at the Department of Computer Science and Technology, IIEST Shibpur as their expertise and dedication have been instrumental in shaping my academic journey and expanding my knowledge in profound ways.

Finally, I would like to thank my parents Dr. Alak Kumar Datta and Mrs. Jayati Datta, and my friends for their infinite love and support. They encouraged me to pursue my goals and to never give up. I could not have completed this project without them.

Sincerely,

Oishika Datta
Intern, Bureau of Indian Standards

EXECUTIVE SUMMARY

The project titled "Study of Artificial Intelligence Regulations Worldwide to Identify Standardization Opportunities" was approached through a structured methodology encompassing three main phases. Firstly, a comprehensive analysis of global AI regulations was conducted to assess current and upcoming frameworks and their implications on standardization efforts. Secondly, existing and emerging standards in AI were rigorously examined. Finally based on the above data, conclusions were drawn to identify gaps and areas requiring further development. These efforts aimed to provide a foundational understanding of the regulatory landscape and its alignment with evolving technological advancements.

Furthermore, an industrial visit was undertaken to acquire firsthand insights into real-world AI applications and the pressing need for standardization. This engagement with industry experts illuminated critical challenges and opportunities, underscoring the necessity for cohesive standards to foster innovation and ensure ethical deployment of AI technologies.

By bridging regulatory gaps and promoting unified standards, stakeholders can collectively enhance AI's societal impact while mitigating risks associated with its rapid adoption. This holistic approach underscores the project's commitment to fostering a sustainable and inclusive AI ecosystem globally.

TABLE OF CONTENTS

SUBJECT AREA

Artificial Intelligence in the 21st century world - Navigating the balance between the boon and the bane

Artificial Intelligence (AI) stands at the forefront of transforming industries and societies worldwide, promising unprecedented advancements in automation, decision-making, and efficiency. As AI technologies proliferate across sectors such as healthcare, finance, and transportation, the need for robust regulations becomes increasingly urgent. Regulations are essential to ensure ethical AI deployment, safeguarding against potential risks like privacy breaches, algorithmic biases, and societal disruptions.

Implementing effective regulations hinges on the establishment of comprehensive standards that guide AI development, deployment, and usage. Standards play a pivotal role in defining best practices, interoperability requirements, and safety protocols across diverse AI applications. They provide a framework for stakeholders—governments, industries, and researchers—to align technological advancements with ethical considerations and societal needs.

Furthermore, standards facilitate international cooperation, fostering a unified approach to AI governance amidst varying regional policies and practices. They promote transparency, accountability, and trust in AI systems, crucial for public acceptance and regulatory compliance. By adhering to standardized frameworks, organizations can navigate regulatory landscapes more effectively, mitigate legal and operational risks, and accelerate innovation responsibly.

In essence, the convergence of AI with regulatory frameworks and standardized practices not only addresses immediate challenges but also paves the way for a sustainable AI-driven future. Balancing innovation with ethical guidelines ensures that AI technologies contribute positively to global progress while minimizing potential adverse impacts on individuals and societies.

OBJECTIVE:

“STUDY OF ARTIFICIAL INTELLIGENCE REGULATIONS WORLDWIDE TO IDENTIFY STANDARDIZATION OPPORTUNITIES”

METHODOLOGY

To identify standardization opportunities, a methodical approach is crucial. Firstly, conducting an in-depth analysis of regulations from various regions and international bodies allows us to comprehend the diverse regulatory landscapes shaping AI deployment. This step ensures a comprehensive understanding of the regulatory frameworks governing AI technologies globally.

Secondly, examining existing and emerging AI standards provides insights into industry practices and technological requirements. Standards play a pivotal role in ensuring interoperability, reliability, and safety of AI systems across different applications and sectors. Understanding these standards is essential to gauge the current state of technological development and regulatory compliance.

Thirdly, by scrutinizing the gaps between regulations and standards, we can pinpoint critical discrepancies and inconsistencies. These gaps often highlight areas where standardization efforts are urgently needed to align regulatory requirements with industry practices.

This study is thus conducted in three phases:

- Step 1: An analysis of Artificial Intelligence Regulations proposed/enforced by authorities in various regions, countries and international bodies

- Step 2: An overview on Artificial Intelligence Standards

- Step 3: Analyzing Gaps between the Regulations and the Standards in the field of Artificial Intelligence

Phase 1 of 3

An analysis of Artificial Intelligence Regulations proposed/enforced by authorities in various regions, countries and international bodies:

Artificial Intelligence (AI) is rapidly transforming industries and societies around the world. As AI technologies advance, there is a growing need for comprehensive regulations to ensure their ethical and responsible use. Governments and international bodies have recognized the importance of establishing regulatory frameworks to address the challenges and opportunities presented by AI. These regulations aim to promote transparency, fairness, accountability, and the protection of human rights while fostering innovation and economic growth.

We enlist a detailed analysis of the key AI regulations and proposals put forth by various regions, countries, and international bodies. By examining these regulatory initiatives, we can better understand the global landscape of AI governance and the diverse approaches being taken to address the multifaceted issues associated with AI deployment.

Table 1 summarizes the significant AI regulations and proposals, highlighting the regions or bodies responsible, the dates of implementation or proposal, and the key points of each regulation.

Table 1: Summary of Artificial Intelligence Regulations all around the globe

SL NO	REGULATION/PROPOSAL	REGION /BODY	DATE	KEY REQUIREMENTS	SECTION/CLAUSE
1.	EU AI Act [1]	The European Union	Approved by the European Parliament on March 13, 2024	**Safety and Compliance:** AI systems must be safe and comply with existing laws on fundamental rights and Union values.	Section 1.2 (Consistency with existing policy provisions) Section 2.3 (Proportionality)
				Risk Management: High-risk AI systems must comply with specific mandatory requirements related to data quality, documentation, transparency, human oversight, robustness, accuracy, and cybersecurity.	Title III, Chapter 2 (Requirements for High-Risk AI Systems), Article 9 (Risk management system)
				Conformity Assessment: High-risk AI systems must undergo a conformity assessment procedure before being placed on the market or put into service.	Title III, Chapter 3 (Obligations of Providers and Users of High-Risk AI Systems), Article 19 (Conformity assessment)
				Technical documentation: The technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date.	Title III, Chapter 2 (Requirements for High-Risk AI Systems), Article 11 (Technical Documentation)
				Record keeping: High-risk AI systems shall be designed and developed with capabilities enabling the automatic recording of events (‘logs’) while the high-risk AI systems are operating.	Title III, Chapter 2 (Requirements for High-Risk AI Systems), Article 12 (Record Keeping)
				Transparency Obligations: Certain AI systems must inform users when they are interacting with AI, especially in cases involving emotion detection or content generation (e.g., deep fakes).	Title IV (Transparency Obligations for Certain AI Systems), Article 13 (Provision of information to users)

				Post-Market Monitoring: Providers must have a post-market monitoring system in place to address any emerging risks from AI systems that continue to learn after being placed on the market.	Title VIII (Monitoring and Reporting Obligations), Article 67 (Compliant AI systems which present a risk)
				Reporting Obligations: Providers must inform national competent authorities about serious incidents or malfunctions that breach fundamental rights obligations.	Article 16 (Obligations of providers of high-risk AI systems), Article 67 (Compliant AI systems which present a risk)
				Quality Management System: Providers of high-risk AI systems must implement a quality management system to ensure compliance with the regulation.	Article 17 (Quality management system)
				Registration Obligations: Providers must register their high-risk AI systems with national competent authorities before placing them on the market.	Article 51 (Registration obligations)
				Penalties for Non-Compliance: Member States must establish effective, proportionate, and dissuasive penalties for infringements of the regulation.	Section: Article 84 (Penalties for infringements)
2.	A pro-innovation approach to AI regulation - March 2023 [2]	The United Kingdom	March 2023 - White Paper consultation on regulating Artificial Intelligence (AI) by DSIT 6th February 2024 – Response by the UK Government	**Safety, security and robustness:** AI systems should function in a robust, secure and safe way throughout the AI life cycle, and risks should be continually identified, assessed and managed.	3.2.3 A principles-based approach
				Appropriate transparency and explainability: An appropriate level of transparency and explainability will mean that regulators have sufficient information about AI systems and their associated inputs and outputs to give	3.2.3 A principles-based approach

				meaningful effect to the other principles.	
				Fairness: AI systems should not undermine the legal rights of individuals or organizations, discriminate unfairly against individuals or create unfair market outcomes. Actors involved in all stages of the AI life cycle should consider definitions of fairness that are appropriate to a system's use, outcomes and the application of relevant law.	3.2.3 A principles-based approach
				Accountability and governance: Governance measures should be in place to ensure effective oversight of the supply and use of AI systems, with clear lines of accountability established across the AI life cycle.	3.2.3 A principles-based approach
				Contestability and redress: Where appropriate, users, impacted third parties and actors in the AI life cycle should be able to contest an AI decision or outcome that is harmful or creates material risk of harm.	3.2.3 A principles-based approach
				Monitoring, assessment and feedback	Box 3.1: Functions required to support implementation of the framework
				Cross-sectoral risk assessment	3.2.3 A principles-based approach
3.	Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence [3]	The United States of America	Published by the White House on October 30, 2023	**Safe and secure AI:** Appropriate safeguards against fraud, and addressing AI systems' most pressing security risks.	Sec. 2. Policy and Principles (a), (e) 4.1. Developing Guidelines, Standards, and Best Practices for AI Safety (i) 4.2. Ensuring Safe and Reliable AI. (a)
				Responsible innovation, and development of AI	Sec. 2. Policy and Principles (b), (c)
				Protection of privacy and civil liberties, promotion of social equity: Removal of unintended bias, discrimination, infringements on privacy, and other harms from AI.	Sec. 2. Policy and Principles. (d), (e), (f) Sec. 7. Advancing Equity and Civil Rights.

				Assessment of High-risk and potentially harmful AI systems: Launching an initiative to create guidance and benchmarks for evaluating and auditing AI capabilities, with a focus on capabilities through which AI could cause harm, such as in the areas of cybersecurity and biosecurity.	4.1. Developing Guidelines, Standards, and Best Practices for AI Safety (i) (C) Managing AI in Critical Infrastructure and in Cybersecurity. (a) To ensure the protection of critical infrastructure, the following actions shall be taken
				Responsible Governance of and by AI	Sec. 2. Policy and Principles. (g).
				Identifying and labeling synthetic content produced by AI systems and establishing the authenticity of digital content: (i) authenticating content and tracking its provenance; (ii) labeling synthetic content, such as using watermarking; (iii) detecting synthetic content; (iv) preventing generative AI from producing child sexual abuse material or producing non-consensual intimate imagery of real individuals (to include intimate digital depictions of the body or body parts of an identifiable individual); (v) testing software used for the above purposes; and (vi) auditing and maintaining synthetic content.	4.5. Reducing the Risks Posed by Synthetic Content.

4.	The Artificial Intelligence and Data Act (AIDA) [4]	Canada	2024 (initially introduced as part of Bill C-27, the Digital Charter Implementation Act, 2022)	**Separation of High-impact AI systems from regular 'harmless' AI systems:** The following key factors are to be examined in determining which AI systems would be considered to be high-impact: (i) Evidence of risks of harm to health and safety, or a risk of adverse impact on human rights, based on both the intended purpose and potential unintended consequences; (ii) The severity of potential harms; (iii) The scale of use; (iv) The nature of harms or adverse impacts that have already taken place; (v) The extent to which for practical or legal reasons it is not reasonably possible to opt-out from that system; (vi) Imbalances of economic or social circumstances, or age of impacted persons; and (vii) The degree to which the risks are adequately regulated under another law.	*"High-impact AI systems: considerations and systems of interest"*
				Human Oversight & Monitoring: People managing the operations of the system should be able to exercise meaningful oversight, aided by crucial measurement and assessment of high-impact AI systems and their output.	*"Regulatory requirements"*

				Transparency of AI systems: Providing the public with appropriate information about how high-impact AI systems are being used.	*"Regulatory requirements"*
				Fairness and Equity: High-impact AI systems must be aware of potential discriminatory outcomes.	*"Regulatory requirements"*
				Safety: High-impact AI systems must be proactively assessed to identify harms that could result from use of the system, including through reasonably foreseeable misuse.	*"Regulatory requirements"*
				Accountability: Organizations must put in place governance mechanisms needed to ensure compliance with all legal obligations of high-impact AI systems in the context in which they will be used.	*"Regulatory requirements"*
				Validity & Robustness: A high-impact AI system must perform consistently with intended objectives, while being stable and resilient in a variety of circumstances.	*"Regulatory requirements"*
5.	New Generation Artificial Intelligence Development Plan [5]	China	2017	**Traceability and accountability:** Establish a traceability and accountability system, and clarify the main body of AI and related rights, obligations, and responsibilities.	V. Guarantee measures (1) Develop laws, regulations, and ethical norms that promote the development of AI
				Development of a new-generation AI theory and technology system: Special focus on emerging technology such as deep learning, cross-domain integration, man-machine collaboration, swarm intelligence, autonomous control.	(2) The Basic Principles (3) Strategic Objectives

6.	Artificial Intelligence (AI) Ethics Principles Framework [6]	Australia	7 November 2019	**Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment.	Principles in detail: Human, social and environmental wellbeing
				Human-centred values: AI systems should respect human rights, diversity, and the autonomy of individuals.	Principles in detail: Human-centred values
				Fairness: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.	Principles in detail: Fairness
				Privacy protection and security: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.	Principles in detail: Privacy protection and security
				Reliability and safety: AI systems should reliably operate in accordance with their intended purpose.	Principles in detail: Reliability and safety
				Transparency and explainability: There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.	Principles in detail: Transparency and explainability

				Contestability: When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.	Principles in detail: Contestability
				Accountability: People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.	Principles in detail: Accountability
7.	Model Artificial Intelligence Governance Framework (Second Edition)[7]	Singapore	Jan 21, 2020	**Risk management:** Authorities aim to build stakeholder confidence in AI through organisations' responsible use of AI to manage different risks in AI deployment	Objectives 2.3 a.
				Data management and protection: Demonstrate reasonable efforts to align internal policies, structures and processes with relevant accountability-based practices in data management and protection.	Objectives 2.3 b.
				Explainability, transparency and fairness: Organisations using AI in decision-making should ensure that the decision-making process is explainable, transparent and fair	Guiding Principles 2.7 a.
				Human-centric AI: The protection of the interests of human beings, including their well-being and safety, should be the primary considerations in the design, development and deployment of AI	Guiding Principles 2.7 b.
8.	National Strategy for Artificial Intelligence [8]	India	June, 2018	**Fairness:** Acknowledging that the existing data may have biases, which may have been reinforced over time, and thus, developing better, bias-free models.	Ethics, Privacy, Security and Artificial Intelligence: Ethics and AI: Fairness / tackling the biases AI

				Transparency: "Opening the Black Box" for AI systems should not aim towards opening of code or technical disclosure – few clients of AI solutions would be sophisticated AI experts - but should rather aim at "explainability".	Ethics, Privacy, Security and Artificial Intelligence: Ethics and AI: Transparency / opening the "Black Box"
				Privacy: Ensure privacy of user data based on the 7-core principles of data protection and privacy – informed consent, technology agnosticism, data controller accountability, data minimisation, holistic application, deterrent penalties and structured enforcement	Ethics, Privacy, Security and Artificial Intelligence: Privacy and AI
				Security: a. Negligence test for damages caused by AI software, as opposed to strict liability. b. As an extension of the negligence test, safe harbours need to be formulated. c. Framework for apportionment of damages need to be developed. d. Actual harm requirements policy may be followed.	Ethics, Privacy, Security and Artificial Intelligence: Security in AI
9.	National Guidelines for Artificial Intelligence (AI) Ethics [9]	South Korea	December 23, 2020	**Safeguarding Human Rights:** AI should not be developed or utilized in a way that violates human rights and freedom.	3. Ten Key Requirements
				Protection of Privacy: The privacy of individuals should be protected throughout the entire process of AI development and utilization.	3. Ten Key Requirements
				Respect for Diversity: Throughout every stage of AI development and utilization, the diversity and representativeness of the AI users should be ensured, and bias and discrimination based on personal characteristics, such as gender, age, disability, region, race, religion, and nationality, should be minimized.	3. Ten Key Requirements
				Prevention of Harm: AI should not be used for the	3. Ten Key Requirements

				purpose of inflicting direct or indirect harm on humans.	
				Public Good: AI should be utilized not only for the pursuit of personal happiness but also for the public good of society and the common benefit of humanity.	3. Ten Key Requirements
				Solidarity: AI should be utilized in a way that helps maintain solidarity among various groups and takes into account the needs of future generations.	3. Ten Key Requirements
				Data Management: Data, such as personal information, should not be used for purposes other than its intended use.	3. Ten Key Requirements
				Accountability: Responsible parties should be clearly defined during the process of AI development and utilization to minimize potential damage.	3. Ten Key Requirements
				Safety: Throughout the entire process of AI development and utilization, efforts should be made to prevent potential risks and ensure safety.	3. Ten Key Requirements
				Transparency: Efforts should be made to improve the transparency and explainability of AI to a level suitable for the use cases of the AI system.	3. Ten Key Requirements
10.	Social Principles of Human-Centric AI Governance Guidelines for Implementation of AI Principles [10]	Japan	January 15, 2021	**Human-Centric:** The utilization of AI must not infringe upon the fundamental human rights guaranteed by the Constitution and international standards.	4.1 Social Principles of AI
				Education/Literacy: Policy makers and managers of businesses involved in AI must have an accurate understanding of AI, knowledge and ethics permitting appropriate use of AI in society. Furthermore, AI users should have a general understanding of AI and should acquire sufficient education to use it properly.	4.1 Social Principles of AI
				Privacy Protection: Careful discretion may be required while handling personal data in accordance with the level of	4.1 Social Principles of AI

				importance and sensitivity of the data.	
				Ensuring Security: The set of risks to security posed by AI should be appropriately addressed.	4.1 Social Principles of AI
				Fair Competition: A fair competitive environment must be maintained in order to create new businesses and services, to maintain sustainable economic growth, and to present solutions to social challenges.	4.1 Social Principles of AI
				Fairness, Accountability, and Transparency: It is necessary to ensure fairness and transparency in decision-making, appropriate accountability for the results, and trust in the technology, so that people who use AI are not subject to undue discrimination.	4.1 Social Principles of AI 4.1 Social Principles of AI
11.	Israel's Policy on Artificial Intelligence Regulation and Ethics [11]	Israel	2023	**A Risk-Based approach:** AI regulation should be adapted to the risks posed by the type of technology, weighted against the potential benefits and risk mitigation measures that are applied in the context of the specific use being regulated.	1. Establishing a governmental policy framework for AI regulation
				Human-centric AI: The development and use of an AI system should respect the rule of law, fundamental rights and public interests, and in particular, it should preserve human dignity and the right to privacy.	2. Adopting a common set of ethical AI principles
				Equality and non-discrimination: Consideration should be given to risks of biases and discrimination against individuals or groups, while bearing in mind AI's potential to promote equality.	2. Adopting a common set of ethical AI principles
				Transparency and explainability: To the extent possible and in appropriate cases, individuals should be:	2. Adopting a common set of ethical AI principles

				(1) informed that they are interacting with an AI system, (2) notified if an AI system is being used to make recommendations or decisions involving them, and (3) provided with an understandable explanation of an AI-based recommendation or decision involving them.	
				Reliability, robustness, security and safety: Suitable measures should be taken, in accordance with generally accepted professional risk management standards, in order to mitigate potential safety and cyber-related risks throughout the lifecycle of AI systems.	2. Adopting a common set of ethical AI principles
				Accountability: Developers, operators and users of AI should be accountable for the proper functioning of AI systems and for the implementation of the other ethical principles in their operation.	2. Adopting a common set of ethical AI principles
12.	Bill 2338/2023 [12]	Brazil	May 2023	**Bias mitigation**	CHAPTER I Art. 3 IV - non - discrimination V - fairness, equity and inclusion Art. 4 VI - discrimination VII - indirect discrimination
				Transparency	CHAPTER I Art. 3 VI - transparency, explainability, intelligibility and auditability
				Reliability and Robustness of AI	CHAPTER I Art. 3 VII - reliability and robustness of artificial intelligence systems and information security;
				Traceability of AI systems	CHAPTER I Art. 3

					IX - traceability of decisions during the life cycle of artificial intelligence systems as a means of accountability and attribution of responsibilities to a natural or legal person;
				Accountability	CHAPTER I Art. 3 X - accountability, liability and full reparation of damages;
				Safe and secure AI	CHAPTER I Art. 3 XI - prevention, precaution and mitigation of systemic risks derived from intentional or unintentional uses and from unforeseen effects of artificial intelligence systems; and XII - non-maleficence and proportionality between the methods employed and the determined and legitimate purposes of artificial intelligence systems.
				Risk-based approach: It shall be the responsibility of the competent authority to regulate artificial intelligence systems of excessive risk. It shall be the responsibility of the competent authority to update the list of excessive or high-risk artificial intelligence systems, identifying new cases, based on certain criteria (Art 18)	CHAPTER III RISK CATEGORIZATION Section II Excessive Risk Section III High Risk
13.	Document A/78/L.49 Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development [13]	The UN General Assembly	11th March 2024	**Safe, secure and trustworthy AI systems:**	2. “Resolves to promote safe, secure and trustworthy artificial intelligence …… shared global challenges, particularly for developing countries.”
				Transparency	6. j. “Safeguarding privacy and the protection …. artificial intelligence systems”
				Diversity, equity and inclusion	9. “Encourages the private sector to adhere to

					applicable international and domestic laws ..… for developing countries”
14.	Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law [14]	The European Commission	9th May, 2024	**Risk and Impact Management:** Each Party shall adopt or maintain measures for the identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems by considering actual and potential impacts to human rights, democracy and the rule of law.	Chapter V – Assessment and mitigation of risks and adverse impacts Article 16 – Risk and impact management framework
				Human dignity and individual autonomy: Each Party shall adopt or maintain measures to respect human dignity and individual autonomy in relation to activities within the lifecycle of artificial intelligence systems.	Chapter III – Principles related to activities within the lifecycle of artificial intelligence systems Article 7
				Transparency and oversight: Each Party shall adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place in respect of activities within the lifecycle of artificial intelligence systems, including with regard to the identification of content generated by artificial intelligence systems.	Chapter III – Principles related to activities within the lifecycle of artificial intelligence systems Article 8
				Accountability and responsibility: Implementing measures to ensure accountability and responsibility for adverse impacts on human rights, democracy and the rule of law resulting from activities within the lifecycle of artificial intelligence systems.	Chapter III – Principles related to activities within the lifecycle of artificial intelligence systems Article 9
				Equality and non-discrimination: Adoption of measures to ensure that activities within the lifecycle of artificial intelligence systems respect equality, including gender equality, and the prohibition of discrimination, as provided under applicable international and domestic law.	Chapter III – Principles related to activities within the lifecycle of artificial intelligence systems Article 10

				Privacy and personal data protection: Measures to ensure protection of privacy rights of individuals and their personal data.	Chapter III – Principles related to activities within the lifecycle of artificial intelligence systems Article 11
				Reliability: Promoting the reliability of artificial intelligence systems and trust in their outputs, which could include requirements related to adequate quality and security throughout the lifecycle of artificial intelligence systems	Chapter III – Principles related to activities within the lifecycle of artificial intelligence systems Article 12
15.	Digital Services Act [15]	The European Union	23rd April 2022.	**Mitigation of Disinformation:** Platforms must implement measures to combat the spread of misinformation and deep-fakes, leveraging AI to detect and manage fake news while being transparent about these processes	(61) "Action against illegal content …. providers of online platforms."
				Transparency and accountability: To ensure an adequate level of transparency and accountability, providers of intermediary services should make publicly available an annual report in a machine-readable format, in accordance with the harmonised requirements contained in the Regulation	(49) "To ensure an adequate level of …. of this Regulation."

This analysis will serve as the baseline for the gap analysis between standards (both published and in development) and the regulations, which is the ultimate aim of the project. By identifying discrepancies and areas of alignment between current standards and regulatory requirements, the project aims to provide insights and recommendations for bridging these gaps.

This concludes the first phase of the project.

Phase 2 of 3

An overview on Artificial Intelligence Standards:

As Artificial Intelligence (AI) continues to evolve and permeate various sectors, establishing robust standards becomes crucial to ensure the technology's ethical and effective deployment. These standards provide guidelines and best practices that foster consistency, reliability, and safety in AI applications. In the context of our project, this analysis represents the second step, following our initial examination of AI regulations.

In this section, we will delve into the various AI standards, both those that are already published and those currently under development. By evaluating these standards, we aim to understand the foundational principles and technical specifications that guide AI development and implementation. This analysis will set the stage for a comprehensive gap analysis, where we will compare these standards with existing and upcoming regulations to identify areas of alignment and discrepancies.

The technical committee LITD 30 at the Bureau of Indian Standards deals with standardization in the field of Artificial intelligence.

Below we list out the published and developing standards on Artificial Intelligence by the ISO/IEC JTC 1/SC 42 subcommittee of the Joint Technical Committee 1 (JTC 1) of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). This subcommittee is focused on the standardization of artificial intelligence (AI), and acts as the Liaison for LITD 30.

PUBLISHED STANDARDS

1. **ISO/IEC 5338: Information technology - Artificial intelligence - AI system life cycle processes**

   The document defines processes for the life cycle of AI systems, focusing on machine learning and heuristic systems. It integrates AI-specific processes with traditional software and system life cycle models, emphasizing acquisition, management, execution, and improvement stages. Key concepts include AI system definition, life cycle model management, infrastructure and quality management, risk and configuration management, and continuous validation and maintenance processes. These processes aim to enhance efficiency, adoption, and stakeholder understanding in developing and managing AI systems across various applications.

2. **ISO/IEC 5339: Information technology - Artificial intelligence - Guidance for AI applications**

   This document provides comprehensive guidance for developing and applying AI applications, detailing the context, stakeholders, processes, and lifecycle stages involved. It emphasizes understanding AI's functional and non-functional characteristics, including trustworthiness, ethics, and societal impacts, to ensure responsible and effective AI deployment. The AI application framework incorporates the perspectives of various stakeholders - focusing on their roles, responsibilities, and the implications of AI deployment.

3. **ISO/IEC 5392: Information technology - Artificial intelligence - Reference architecture of knowledge engineering**

   The document outlines a reference architecture for knowledge engineering (KE) in AI, detailing KE roles, activities, constructional layers, and components. It emphasizes the integration of human knowledge into machine-understandable formats across industries like finance, healthcare, and transportation. Key technologies include knowledge representation, acquisition, fusion, and visualization. The document also defines a common vocabulary and aims to guide the construction and collaboration of KE systems.

4. **ISO/IEC TR 5469: Artificial intelligence - Functional safety and AI systems**

   It addresses the application of artificial intelligence (AI) technologies in safety-related systems, focusing on their properties, risk factors, and functional safety methods. It highlights the challenges in specifying, designing, and verifying AI systems, particularly machine learning (ML), due to their complex and often non-transparent nature. The report outlines a three-stage realization principle for using AI in safety functions, discusses properties and risk factors such as transparency, explainability, and resilience to adversarial inputs, and offers solutions for verification and validation, including control and mitigation measures. Additionally, it explores

the integration of AI within existing safety standards and provides methodological guidance for developing safe AI systems.

5. **IS/ISO/IEC 8183: Information technology - Artificial intelligence - Data life cycle framework**

   The document outlines a comprehensive data life cycle framework for artificial intelligence (AI) systems, spanning from idea conception to system and data decommissioning. It defines stages such as business requirements, data planning, acquisition, preparation, model building, system deployment, operation, and eventual decommissioning. Each stage involves specific actions like data acquisition, cleaning, feature engineering, and model validation, ensuring alignment with business goals, ethical standards, and regulatory requirements throughout the AI system's life-cycle. The framework aims to guide organizations in effectively managing data for AI development and deployment.

6. **ISO/IEC TS 8200: Information technology - Artificial intelligence - Controllability of automated artificial intelligence systems**

   The document provides a framework outlining principles, characteristics, and approaches for enhancing AI system controllability. It addresses key areas such as state observability, state transition, control transfer processes, reaction to uncertainty during control transfers, and verification and validation methods. Emphasizing the importance of control over AI systems to ensure safety and reliability, the document discusses the design and implementation considerations for controllability, including the cost and collaborative aspects of control. It also covers requirements specific to continuous learning systems and safety-critical AI system design, guiding organizations in developing and using AI systems throughout their lifecycle.

7. **IS/ISO/IEC 22989: Information technology - Artificial intelligence - Artificial intelligence concepts and terminology**

   The standard covers AI concepts such as machine learning, cognitive computing, and genetic algorithms, and discusses important aspects like AI trustworthiness, robustness, reliability, and explainability. The document emphasizes the lifecycle of AI systems, detailing stages from inception to retirement, and highlights the functional overview of AI systems, including data processing, learning, prediction, and decision-making. Additionally, it explores the AI ecosystem, including big data, cloud computing, and resource pools, and examines various AI fields and applications, such as computer vision, natural language processing, and fraud detection. This comprehensive guide is intended for a wide audience, including both experts and non-practitioners, to support the development and communication of AI standards.

8. **IS/ISO/IEC 23053: Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)**

   This document establishes a comprehensive framework for AI systems using machine learning

(ML), applicable to organizations of all types and sizes. It provides a detailed description of AI system components, their functions, and the AI ecosystem, with a focus on ML methodologies including deep learning. Key sections include an overview of ML tasks such as regression, classification, and clustering, as well as detailed descriptions of ML models, data handling, tools, optimization methods, and evaluation metrics. The document also outlines various ML approaches like supervised, unsupervised, semi-supervised, self-supervised, reinforcement learning, and transfer learning. Additionally, it presents a structured ML pipeline covering data acquisition, preparation, modeling, verification, validation, and deployment, along with an example process for practical implementation.

9. **ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management**

   The document provides guidance on risk management for organizations that develop, produce, deploy, or use AI systems, aligning with ISO 31000:2018 principles. It details how AI-specific risk management can enhance performance, support innovation, and help achieve objectives. Divided into three main parts, it covers principles of risk management (Clause 4), the framework for integrating risk management into organizational functions (Clause 5), and processes for implementing AI risk management (Clause 6). The document emphasizes the importance of understanding organizational context, leadership commitment, resource allocation, and continuous improvement. This comprehensive guide can be customized to fit any organization's context, ensuring effective AI risk management.

10. **ISO/IEC TR 24027: Information technology - Artificial intelligence (AI) - Bias in AI systems and AI aided decision making**

    The Technical Report provides comprehensive guidance on addressing bias in AI systems, particularly in the context of AI-aided decision-making. It covers all phases of the AI system lifecycle, emphasizing the identification, assessment, and treatment of bias-related vulnerabilities. The document outlines various sources of bias, including human cognitive biases, data biases, and biases introduced by engineering decisions. It offers measurement techniques and methods for evaluating bias and fairness through metrics like confusion matrices, equalized odds, and demographic parity. Furthermore, the standard provides strategies for treating unwanted bias across the AI system lifecycle, from inception and design to verification, validation, and deployment. Transparency tools and continuous monitoring are also highlighted to ensure ongoing mitigation of bias in AI systems.

11. **IS/ISO/IEC TR 24028: Information technology - Artificial intelligence - Overview of trustworthiness in artificial intelligence**

    The standard focuses on trustworthiness in AI systems, addressing transparency, explainability, and controllability as essential factors to establish trust. It identifies engineering pitfalls and typical threats to AI systems, emphasizing mitigation techniques. The document also outlines approaches to assess and achieve availability, resiliency, reliability, accuracy, safety, security, and

privacy in AI systems. Key concerns such as bias, unpredictability, and system opaqueness are highlighted, with strategies provided for their mitigation throughout the AI system lifecycle. Stakeholder considerations, including responsibility, accountability, and governance, are central to ensuring trust in AI applications. The document concludes by emphasizing the importance of systematic risk-based approaches and transparent communication to enhance trust in AI systems across various stakeholders and societal contexts.

12. **IS/ISO/IEC TR 24368: Information technology - Artificial intelligence - Overview of ethical and societal concerns**

    It provides a comprehensive overview of ethical and societal concerns related to artificial intelligence (AI). It addresses fundamental sources of ethical issues such as privacy breaches, biased data, and opaque decision-making processes. The document outlines various ethical frameworks including virtue ethics, utilitarianism, and deontology, and discusses human rights practices relevant to AI applications. Key themes and principles covered include accountability, fairness, transparency, privacy, safety, and human control of technology. It also offers practical examples and considerations for organizations to align their processes with ethical AI principles, emphasizing the importance of ethical reviews, international norms, and sustainability in AI development and deployment.

13. **ISO/IEC TS 25058: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Guidance for quality evaluation of artificial intelligence (AI) systems**

    The document provides a specialized AI system quality model based on SQuaRE, tailored for assessing AI systems. It covers key quality characteristics such as functional completeness, correctness, adaptability, performance efficiency, compatibility, usability, and transparency. The standard emphasizes using appropriate metrics and testing methods like functional and metamorphic testing to ensure robust evaluations. It aims to guide organizations in meeting ISO/IEC standards for reliability, security, maintainability, and portability, while addressing ethical considerations and environmental impacts, promoting responsible AI development and deployment.

14. **ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems**

    The document defines a comprehensive quality model tailored specifically for AI systems, building upon the SQuaRE framework. It introduces new and modified characteristics to address the unique challenges of AI, such as functional adaptability, robustness, transparency, and intervenability. The document emphasizes the importance of evaluating AI system quality from both product and quality in use perspectives, ensuring usability, reliability, and societal responsibility. It provides guidance on assessing characteristics like user controllability and ethical risk mitigation, aiming to enhance transparency, accountability, and safety throughout the AI system lifecycle. Overall, ISO/IEC 25059 aims to standardize the evaluation and

measurement of AI system quality to meet evolving technological demands and ethical considerations.

15. ISO/IEC 42001: Information technology - Artificial intelligence - Management system

It specifies requirements and guidance for establishing, implementing, maintaining, and improving an AI management system within organizations that develop or utilize AI systems. It aims to help organizations responsibly navigate the complexities of AI technologies by addressing specific challenges such as transparency, continuous learning, and ethical considerations. The standard outlines processes for risk management, lifecycle management, and quality assurance tailored to AI systems, ensuring they align with organizational objectives and stakeholder expectations. It emphasizes leadership commitment, AI policy development, and integration with existing management structures to foster accountability and operational excellence. By adopting a harmonized structure with other management system standards, it promotes consistency and facilitates effective implementation across diverse AI applications and sectors.

16. IS/ISO/IEC TS 4213: Information technology - Artificial intelligence - Assessment of machine learning classification performance

The document outlines methodologies for assessing the classification performance of machine learning models, emphasizing a generalized process involving task determination, metric specification, evaluation, data collection, and result generation. It highlights the importance of addressing data bias, preprocessing, cross-validation, and preventing information leakage while considering computational complexity, efficiency, and energy consumption during evaluation.

17. ISO/IEC TR 17903: Information technology - Artificial intelligence - Overview of machine learning computing devices

This document surveys machine learning (ML) computing devices, covering terminology, characteristics, and performance optimization approaches. It highlights processing, computing, device infrastructure, and service concepts, emphasizing the importance of data types, ML operators, memory access, scheduling, topologies, streams, buffering, and caching mechanisms. Performance optimization focuses on computational graph optimization, ML operator optimization, and system efficiency, with key measures including time consumption, throughput, and power consumption. The information is relevant for organizations of all types and sizes, aiming to enhance the effectiveness and efficiency of ML computing devices within AI systems.

18. ISO/IEC 20546: Information technology - Big data - Overview and vocabulary

This document provides an overview and vocabulary for big data, focusing on key data characteristics (volume, velocity, variety, and variability) and processing characteristics (data science, volatility, veracity, visualization, and scaling). It addresses the significance of structured

and unstructured data, distributed file systems, distributed data processing, and non-relational (NoSQL) databases, emphasizing horizontal scaling and the handling of diverse data types. The information aims to enhance communication and understanding in the field of big data across various technical areas.

19. **ISO/IEC TR 20547-1: Information technology - Big data reference architecture - Part 1: Framework and application process**

    Provides a framework for the Big Data Reference Architecture (BDRA) and a process for applying it to specific problem domains. It highlights key big data characteristics (volume, velocity, variety, variability) and addresses stakeholders' concerns, emphasizing security, privacy, and interoperability.

    The BDRA offers a structure for big data systems, defining roles, activities, functional components, and views to ensure scalable and efficient data processing. It also outlines steps for identifying stakeholders, mapping concerns to roles, detailing activities, defining functional components, and validating the architecture.

20. **ISO/IEC TR 20547-2: Information technology - Big data reference architecture - Part 2: Use cases and derived requirements**

    Outlines examples of big data use cases across various domains, detailing their application areas and technical considerations. It covers sectors such as government, commercial, defense, healthcare, deep learning, research ecosystems, astronomy, environmental science, and energy. Each use case is described in terms of its current solutions, big data characteristics, and related issues.
    The document aims to provide a comprehensive template and framework for understanding and applying big data solutions in diverse contexts

21. **ISO/IEC 20547-3: Information technology - Big data reference architecture - Part 3: Reference architecture**

    The document specifies the Big Data Reference Architecture (BDRA), defining user and functional views. The user view includes roles, sub-roles, and activities within a big data ecosystem, while the functional view outlines architectural layers and functional components necessary for implementing these activities. Key objectives of the BDRA are to provide a common language, encourage adherence to standards, ensure consistency, facilitate understanding of big data operations, and support technical references for stakeholders. It emphasizes the relationships between components and cross-cutting aspects like security, privacy, and data governance.

22. **ISO/IEC TR 20547-5: Information technology - Big data reference architecture - Part 5: Standards roadmap**

The document provides an overview of existing and developing standards relevant to big data, emphasizing the need for a cohesive approach to integrate these standards into the Big Data Reference Architecture (BDRA). It identifies key standards organizations and industry consortia, such as ISO, IEEE, and W3C, and lists relevant standards across various big data functional layers. The document also highlights potential gaps in current big data standardization, including areas like metadata specifications, query languages, and security controls, and suggests pathways for addressing these gaps through the evolution of existing standards and development of new theories and best practices.

23. **IS/ISO/IEC TR 24029-1: Artificial Intelligence (AI) - Assessment of the robustness of neural networks - Part 1: Overview**

    Overview of methods to assess the robustness of neural networks, focusing on **statistical, formal, and empirical approaches.**
    A typical workflow includes stating robustness goals, planning and conducting tests, analyzing outcomes, and interpreting results.
    Statistical methods involve performance metrics, formal methods use mathematical proofs, and empirical methods rely on human judgment and field trials to evaluate robustness.

24. **IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods**
    States the methodologies for using formal methods to assess the robustness of neural networks.Robustness assessment focuses on **stability, sensitivity, relevance, and reachability** properties across different domains and input data types.
    Various formal methods, including solvers, abstract interpretation, reachability analysis, and model checking, are discussed for evaluating robustness throughout the neural network lifecycle—from design and development to deployment and operation monitoring.

25. **IS/ISO/IEC TR 24030: Information technology - Artificial intelligence (AI) - Use cases**

    The document provides a collection of representative AI application use cases across various domains covering general application domains, deployment models, and specific examples of AI applications.
    Detailed use cases are presented for agriculture, digital marketing, education, energy, fintech, healthcare, home/service robotics, ICT, legal, logistics, maintenance and support, manufacturing, media and entertainment, mobility, public sector, retail, security, social infrastructure, transportation, work and life, and others.
    The document also includes guidance on submitting use cases, acceptable sources, properties, basic statistics, societal concerns, and opportunities for standardization.

26. **IS/ISO/IEC TR 24372: Information technology - Artificial intelligence (AI) - Overview of computational approaches for AI systems**

A comprehensive overview of computational approaches in AI systems, detailing their main characteristics, algorithms, and approaches. It covers both knowledge-driven and data-driven methods, including machine learning and metaheuristics. Key topics include knowledge engineering, logic and reasoning (inductive, deductive, Bayesian), various machine learning models (decision trees, neural networks, GANs), and metaheuristic algorithms like genetic algorithms. The document emphasizes the diverse computational foundations and applications of AI systems across different domains and use cases

27. **IS/ISO/IEC 24668: Information technology - Artificial intelligence - Process management framework for big data analytics**

    The document outlines a framework for implementing big data analytics (BDA) across diverse organizational functions. It introduces a Process Reference Model (PRM) and Process Assessment Model (PAM) structured around five key process categories: organization stakeholder, competency development, data management, analytics development, and technology integration. The framework aims to optimize BDA processes by providing process descriptions, performance indicators, and capability assessments aligned with international standards, facilitating improved decision-making, competitive advantages, and operational efficiencies through BDA automation and enhancement.

28. **IS/ISO/IEC 38507: Information technology - Governance of IT - Governance implications of the use of artificial intelligence by organizations**

    Provides guidance to governing bodies of organizations on effectively governing the use of Artificial Intelligence (AI). It emphasizes the role of governance in managing AI's opportunities, risks, and responsibilities, focusing on human oversight rather than technical aspects. Key topics include understanding AI's governance implications, maintaining accountability, distinguishing AI systems from other technologies, and implementing policies for decision-making, data use, culture, compliance, and risk management. Applicable to all types and sizes of organizations, this document aims to ensure the effective, efficient, and ethical use of AI across various sectors.

STANDARDS UNDER DEVELOPMENT

1. **ISO/IEC AWI TS 22443: Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations**

   The document proposes guidance for organizations to identify and address societal concerns and ethical considerations throughout the life cycle of AI systems, aiming to enhance trustworthiness and compliance with emerging regulations. It covers diverse impacts such as autonomy, privacy, accountability, and effects on sectors like healthcare and education. Stakeholders like industry, SMEs, governments, consumers, and academia are expected to benefit from improved risk management and innovation in AI deployment, fostering dialogue and ensuring alignment with international standards and Sustainable Development Goals.

2. **ISO/IEC AWI 25029: Artificial intelligence - AI-enhanced nudging**

   This standard aims to define and guide the implementation of AI-enhanced nudging mechanisms, ensuring they align with existing AI standards. It provides organizations with guidelines for responsible design and management, emphasizing the protection of individual free will while supporting diverse stakeholders such as consumers, workers, and NGOs. The focus includes balancing general and sector-specific standards, lifecycle management, criteria for evaluation, and managing residual risks in deploying AI-enhanced nudging systems.

3. **ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems**

   This proposal seeks to update ISO/IEC 25059 to include quality models specifically tailored for AI systems and services. It addresses the unique challenges posed by AI, such as adaptability, learning capacity, and reliance on data quality, necessitating a refined quality assessment framework. The update aims to align with advancements in related standards, incorporate new topics like explainability and transparency, and provide additional guidance on metrics to enhance evaluation and ensure alignment with societal values.

4. **ISO/IEC 42102 (ed 1): Information technology - Artificial intelligence - Taxonomy of AI system and methods and capabilities**

   This document proposes a taxonomy for classifying AI systems, aiming to provide a common understanding among stakeholders across various stages of AI system lifecycles. It addresses the need for coherence and interoperability in describing AI applications, supporting development, operation, conformity assessment, and market surveillance. The taxonomy enhances alignment with existing standards and fosters clarity in AI system classification for international and European standardization efforts.

5. **ISO/IEC 42105: Information technology - Artificial intelligence - Guidance for human oversight of AI systems**

   This proposal outlines guidance for human oversight of AI systems, emphasizing the importance of human operators and developers in controlling and monitoring AI throughout its lifecycle. It builds upon ISO/IEC TS 8200 to ensure AI systems remain controllable, providing frameworks for effective policy implementation, information exchange clarity, and appropriate system reactions. The document aims to enhance the safety, trustworthiness, and operational effectiveness of AI systems by defining clear roles, responsibilities, and communication protocols between humans and AI systems.

6. **ISO/IEC 42112: Information technology - Artificial intelligence - Guidance on machine learning model training efficiency optimisation**

   The proposal aims to provide guidance on optimizing machine learning model training efficiency by addressing key characteristics such as communication, storage, and recovery processes. It targets AI providers and producers, offering methodologies like parallelism, resource utilization, and checkpoint mechanisms to enhance training speed and reduce resource consumption. The document aligns with standards on AI efficiency and training processes, ensuring stakeholders can effectively implement and evaluate these optimizations to advance machine learning technology applications.

7. **ISO/IEC FDIS 5259-1: Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 1: Overview, terminology, and examples**

   The document outlines data quality concepts critical for analytics and machine learning (ML), emphasizing the importance of data quality in influencing ML model performance and analytical outcomes. It describes a framework for managing data quality through a defined model, measures, assessment, improvement, and reporting. The six-stage data life cycle model includes data requirements, planning, acquisition, preparation, provisioning, and decommissioning. Key considerations include data security, privacy, governance, and provenance to ensure data integrity and trustworthiness throughout the entire life cycle.

8. **ISO/IEC DIS 5259-2: Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 2: Data quality measures**

   This proposal, part of the ISO/IEC 5259 series, focuses on establishing a standardized approach to manage data quality for analytics and machine learning (ML). It defines a data quality model, characteristics, and measures based on ISO/IEC 25012 and 25024, aiming to ensure high-quality data throughout the data lifecycle. The document provides guidelines for organizations to achieve their data quality objectives, emphasizing the importance of accurate, complete, and current data for reliable and interoperable ML models and analytics processes.

9. **ISO/IEC FDIS 5259-3: Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 3: Data quality management requirements and guidelines**

   This document outlines requirements and guidelines for managing data quality in analytics and machine learning (ML) applications. It emphasizes the importance of a robust data quality management system to ensure the reliability and success of ML models and analytical outcomes. The document covers a broad spectrum of topics including data quality culture, management processes, integration with management systems, documentation, audit, and assessment. It also details life cycle-specific processes such as data motivation, acquisition, preprocessing, augmentation, provisioning, and decommissioning, highlighting the importance of managing data quality throughout its lifecycle stages. Horizontal processes like verification, validation, configuration, change, and risk management are also addressed, ensuring comprehensive coverage for effective data quality management..

10. **ISO/IEC FDIS 5259-4: Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 4: Data quality process framework**

    Establishes a comprehensive data quality process framework for analytics and machine learning (ML). Its primary goal is to ensure consistent and high-quality data management across various ML approaches, including supervised, unsupervised, semi-supervised learning, reinforcement learning, and analytics. The document outlines guidelines for data quality planning, evaluation, improvement, and validation within these contexts. It also includes specific processes for data labeling, participant roles, and considerations unique to each ML approach, thereby aiming to enhance the reliability and effectiveness of AI systems through better data quality management.

11. **ISO/IEC DIS 5259-5: Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 5: Data quality governance framework**
    The document outlines a data quality governance framework for analytics and machine learning (ML), guiding organizations in implementing and overseeing data quality measures and processes. It emphasizes the roles of the governing body and management in establishing strategies, policies, and controls throughout the data lifecycle to ensure high-quality data. The framework applies to organizations of any size and type, aligning with other parts of the ISO/IEC 5259 series.

12. **ISO/IEC DTS 12791.2: Information technology - Artificial intelligence - Treatment of unwanted bias in classification and regression machine learning tasks**

    The document provides guidance on addressing unwanted bias in AI systems, specifically in machine learning tasks involving classification and regression. It outlines steps to mitigate bias throughout the AI system life cycle, applicable to organizations of all sizes. Key sections include identifying stakeholders, defining requirements, and integrating with risk management during inception, as well as techniques like adjusting data and managing risks during design and development. It also covers verification, validation, and continuous monitoring. The document

highlights algorithmic, training, and data techniques to address bias effectively.

13. **ISO/IEC DIS 12792: Information technology - Artificial intelligence - Transparency taxonomy of AI systems**

    The document establishes a taxonomy for transparency in AI systems, aimed at improving trust, accountability, and communication among stakeholders by standardizing terminology and information elements. It addresses various aspects of transparency, including the system's context, internal functioning, and dataset documentation, to aid stakeholders in understanding and evaluating AI systems. The framework is applicable across different industries and regions, providing a foundation for developing specific standards.

14. **ISO/IEC DIS 42005: Information technology - Artificial intelligence - AI system impact assessment**

    The document provides guidance for organizations on performing AI system impact assessments to address potential impacts on individuals and society. It outlines how to integrate these assessments into an organization's AI risk management and overall management system. Key aspects include documenting the process, determining the scope, and analyzing results. This ensures AI systems are trustworthy and transparent, aligning with governance, risk, and conformity assessment practices.

15. **ISO/IEC AWI 24970: Artificial intelligence - AI system logging**

    The document outlines a comprehensive logging system for AI systems, focusing on monitoring, incident investigation, and risk management. It defines logging requirements for various events including AI system inputs, outputs, human interventions, errors, and user interactions. The goal is to enable accurate monitoring of system accuracy, bias, and robustness, and to ensure compliance with legal obligations such as data collection consent. Key contents include guidelines for logging plans, event traceability, compliance measures, and specific event schemas for transaction outcomes, errors, human interventions, user requests, and information transmissions. The document emphasizes the integration of logging with risk management systems to mitigate identified risks effectively.

16. **ISO/IEC CD TR 5259-6: Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 6: Visualization framework for data quality**

    The document outlines a visualization framework tailored for assessing data quality within analytics and machine learning (ML) contexts. It emphasizes the role of visualization in enhancing quality measurement by presenting data quality metrics in a tangible and insightful manner for stakeholders. Visualization aids in data profiling, facilitating cognitive responses during exploratory data analysis, and promoting transparency in ML algorithm operations. The framework aligns with the ISO/IEC 5259 series, aiming to standardize data quality management

practices across organizations. Key components include guidelines for selecting appropriate visualization methods based on quality measures, stakeholder requirements, and stages of the AI system lifecycle. Practical use cases illustrate the framework's application, demonstrating its utility in identifying anomalies, detecting errors, and uncovering relationships within data.

17. **SC42 N1438 NP Outline Guidance Social Ethical concerns**
    Information technology - Artificial intelligence – Guidance for addressing societal concerns and ethical considerations

    The document provides guidelines for organizations to identify and address societal and ethical concerns related to AI systems. It outlines best practices for ensuring ethical behavior, trustworthiness, and the prevention of misuse. The document includes various use cases, highlighting potential benefits and drawbacks of AI applications, and emphasizes the importance of responsible AI development and deployment. It also underscores the need for strategies to mitigate risks such as bias, discrimination, and unintended harmful consequences.

18. **SC42 N1440 NP Outline Human Oversight**
    Artificial intelligence - Guidance for human oversight of AI systems.

    The proposed outline provides guidance on human oversight of AI systems, focusing on the necessary human control and monitoring throughout the AI lifecycle. It extends the framework of ISO/IEC TS 8200, addressing direct and indirect human involvement, monitoring and control, and relationships with other AI system concepts. Key topics include oversight objectives, relevant stakeholders, transparency, explainability, user understanding, and various biases like automation bias.

19. **ISO/IEC AWI 42102: Information technology - Artificial intelligence - Taxonomy of AI system methods and capabilities**

    The document outlines a standard for classifying AI systems, providing a taxonomy based on methods (Traditional AI, Symbolic AI, Machine Learning, Hybrid Learning) and capabilities (Percept, Process, Act, Communicate). It aims to ensure a common understanding among AI stakeholders and is applicable to organizations at any stage of the AI lifecycle. The taxonomy facilitates consistent AI system descriptions and classifications across diverse applications.

20. **ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems**

    Provides a comprehensive framework for achieving explainability in ML models and AI systems. It outlines various approaches such as empirical analysis, post-hoc interpretation (local and global), inherently interpretable components, architecture- and task-driven methods, and data explanation. The standard guides stakeholders—including developers, users, service providers, and policymakers—through considerations for selecting and applying these methods across the

AI system lifecycle to enhance trustworthiness, mitigate bias, comply with regulations, and improve system robustness. It emphasizes the importance of tailored explanations through numeric, visual, textual, and interactive tools, addressing technical constraints and ensuring transparency and usability.

21. **ISO/IEC TR 20226: Information technology - Artificial intelligence - Environmental sustainability aspects of AI systems**

    The working draft outlines the environmental sustainability aspects of AI systems throughout their lifecycle. It covers significant topics such as energy consumption, including the use of power-intensive GPUs and their impact on carbon emissions. Geographic considerations, including location, distribution, and transportation of AI infrastructure, are highlighted. The document addresses water use, cooling methods, carbon footprint, and waste management, emphasizing the potential environmental impacts and providing metrics for measurement. Strategies for reducing these impacts are suggested, including ecosystem, lifecycle, and supply chain approaches, aiming to mitigate environmental degradation caused by AI technologies.

22. **ISO/IEC AWI 23282: Artificial Intelligence - Evaluation methods for accurate natural language processing systems**

    The draft proposal outlines a standard for evaluating natural language processing (NLP) systems, focusing on measuring the quality of system outputs to assess functional suitability. It aims to define specific NLP evaluation metrics, excluding generic AI metrics like precision and recall. Key components include the introduction of task-specific metrics such as BLEU, ROUGE, and WER, detailing their formulas, technical considerations, and implementation requirements. The document emphasizes the importance of metric selection tailored to NLP tasks, beyond classification and regression, with examples illustrating potential pitfalls of using generic metrics. It also discusses human evaluation protocols, qualitative assessments, and considerations for multi-component NLP systems. Requirements on technical resources and test data preparation are highlighted to ensure meaningful evaluation outcomes.

23. **ISO/IEC AWI TS 42112: Information technology - Artificial intelligence - Guidance on machine learning model training efficiency optimisation**

    The document outlines strategies to optimize machine learning (ML) model training efficiency, crucial for reducing time, hardware resources, and costs while maintaining data and model size integrity. It focuses on stakeholders like AI providers and producers, emphasizing the need for faster model deployment and resource-efficient training processes. Key optimization methods include parallelism (data, model, and hybrid), which splits tasks across multiple devices to expedite training. Communication optimization techniques, such as data compression and asynchronous communication, mitigate bottlenecks during parallel operations. Additionally, model checkpointing ensures resilience by periodically saving progress, minimizing downtime due to interruptions. These approaches collectively enhance ML model training efficacy and resource utilization.

Phase 3 of 3

Analyzing Gaps between the Regulations and the Standards in the field of Artificial Intelligence:

This section marks the final phase of our project and constitutes the heart of our report. Having previously examined the regulatory frameworks in the first step and the AI standards in the second step, we now turn our focus to identifying and analyzing the gaps between these regulations and standards. This comprehensive gap analysis aims to document discrepancies and highlight areas that require further standardization.

By systematically comparing the findings from our study of AI regulations with those from our study of AI standards, we will pinpoint specific areas where current standards do not fully align with regulatory requirements. This analysis is crucial for identifying the specific areas where improvements are needed to achieve coherence and integration between regulations and standards in AI. The insights gained from this phase will provide a foundation for developing recommendations to bridge these gaps, thereby fostering a regulatory environment that supports both compliance and innovation in AI technologies.

Our ultimate goal is to contribute to the ongoing dialogue on AI governance and assist stakeholders in creating frameworks that are both effective and forward-looking. Through this analysis, we aim to drive progress towards a more standardized and regulated AI landscape.

After thorough study of the regulations worldwide, and comparison with relevant standards, the particular regulator specifications which are not covered by the current or upcoming standards, and will require the development of new standards/amendments in relevant old standards are listed down in Table 2.

Table 2: AI Regulations with partial or no supporting standards:

SL NO	REGULATION/ PROPOSAL	COUNTRY/ REGION	REQUIREMENT	FULFILLED/ PARTIALLY FULFILLED/ NOT FULFILLED	COMMENTS
1.	EU AI Act [1]	The European Union	Safety and Compliance	**Partially Fulfilled**	**Standards need to be developed classifying AI systems on the different risk-levels prescribed, and analyzed and assessed for potential threats separately.**
			Risk Management	**Partially Fulfilled**	
			Conformity Assessment	**Partially Fulfilled**	
			Technical Documentation	**Partially fulfilled**	**Clause 7.5 Documented information of ISO/IEC 42001: Information technology - Artificial intelligence - Management system covers several aspects of technical documentation pertaining to AI systems in general, However, it does not include the particulars of technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service.**
			Record Keeping	**Partially Fulfilled**	**Although ISO/IEC AWI 24970: Artificial intelligence - AI system logging defines logging requirements for various events including AI system inputs, outputs, human interventions, errors, and user interactions - the concept of automatic logging for high-risk systems, as prescribed in the regulation is absent.**
			Transparency Obligations	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software

					engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems.
			Post-Market Monitoring	Fulfilled	Covered by ISO/IEC 5338: Information technology - Artificial intelligence - AI system life cycle processes, ISO/IEC AWI 24970: Artificial intelligence - AI system logging and C42 N1440 NP Outline Human Oversight Artificial intelligence - Guidance for human oversight of AI systems.
			Reporting Obligations	Fulfilled	Covered by clause 6.7: Recording and Reporting of ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management
			Quality Management System	Fulfilled	Covered by ISO/IEC FDIS 5259-3: Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 3: Data quality management requirements and guidelines
			Registration Obligations	**Partially fulfilled**	**Partially covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management and ISO/IEC 42105: Information technology - Artificial intelligence - Guidance for human oversight of AI systems**
			Penalties for Non-Compliance	Fulfilled	Covered by clause 10.2: Nonconformity and corrective action of ISO/IEC 42001: Information technology - Artificial intelligence - Management system
2.	A pro-innovation approach to AI regulation - March 2023 [2]	The United Kingdom	Safety, security and robustness	Fulfilled	Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.

			Appropriate transparency and explainability	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Fairness and accountability	Fulfilled	
			Governance of AI	Fulfilled	IS/ISO/IEC 38507: Information technology - Governance of IT - Governance implications of the use of artificial intelligence by organizations: This standard addresses the governance aspects of artificial intelligence, emphasizing accountability, liability, transparency, and oversight.
			Contestability and redress	**Partially fulfilled**	**Although clause 10.2: Nonconformity and corrective action of ISO/IEC 42001: Information technology - Artificial intelligence - Management system provides guidance on taking action and dealing with the consequences in cases of non-conformity, further amendments are required to state the guidelines for individuals seeking correction or remedy if they are negatively affected by decisions made by AI systems.**
			Monitoring, assessment and feedback	Fulfilled	Covered by ISO/IEC 5338: Information technology - Artificial intelligence - AI system life cycle processes, ISO/IEC AWI 24970: Artificial intelligence - AI system logging and C42 N1440 NP Outline Human Oversight Artificial intelligence - Guidance for human oversight of AI systems.
			Cross-sectoral risk assessment	**Not fulfilled**	**Standards need to be developed to address the differing levels of risks associated with different sectors.**
3.	Executive Order on the Safe, Secure, and Trustworthy	The United States of America	Safe and secure AI	Fulfilled	ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC

	Development and Use of Artificial Intelligence [3]				23894: Information technology - Artificial intelligence - Guidance on risk management.
			Protection of privacy and civil liberties, promotion of social equity	Fulfilled	Covered by ISO/IEC DTS 12791.2: Information technology - Artificial intelligence - Treatment of unwanted bias in classification and regression machine learning tasks and SC42 N1438 NP Outline Guidance Social Ethical concerns Information technology - Artificial intelligence – Guidance for addressing societal concerns and ethical considerations.
			Assessment of High-risk and potentially harmful AI systems	**Not fulfilled**	**Standards should address the technical specifications for assessing AI models' computing power and data center requirements, ensuring consistency and clarity in regulatory compliance. Additionally, guidelines are necessary for verifying and monitoring adherence to these standards to effectively implement and enforce regulations on AI models potentially used in malicious cyber activities.**
			Responsible Governance of and by AI	Fulfilled	IS/ISO/IEC 38507: Information technology - Governance of IT - Governance implications of the use of artificial intelligence by organizations: This standard addresses the governance aspects of artificial intelligence, emphasizing accountability, liability, transparency, and oversight.
			Identifying and labeling synthetic content produced by AI systems and establishing the authenticity of digital content	**Not fulfilled**	**There are no comprehensive standards specifically aimed at the authenticity and verification of content generated by AI. Standards related to AI do not directly address these issues**
4.	The Artificial Intelligence and Data Act (AIDA) [4]	Canada	Separation of High-impact AI systems from regular ‘harmless’ AI systems	**Not fulfilled**	**Standards need to be developed classifying AI systems on the different risk-levels prescribed, and analyzed and assessed for potential threats separately.**

			Human Oversight & Monitoring	Fulfilled	Covered by ISO/IEC 42105: Information technology - Artificial intelligence - Guidance for human oversight of AI systems
			Transparency of AI systems	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems,
			Fairness and Equity	Fulfilled	
			Accountability	Fulfilled	
			Safety	Fulfilled	Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.
			Validity & Robustness	Fulfilled	
5.	New Generation Artificial Intelligence Development Plan [5]	China	Traceability and accountability	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Development of a new-generation AI theory and technology system	**Not fulfilled**	**Although IS/ISO/IEC 22989 mentions 'swarm intelligence' as a Soft Computing technique (clause 5.7) there are no exclusive standards focussing solely on its specifications and implementation details.**

6.	Artificial Intelligence (AI) Ethics Principles Framework [6]	Australia	Human, societal and environmental wellbeing	Fulfilled	Covered by ISO/IEC DTS 12791.2: Information technology - Artificial intelligence - Treatment of unwanted bias in classification and regression machine learning tasks and SC42 N1438 NP Outline Guidance Social Ethical concerns Information technology - Artificial intelligence – Guidance for addressing societal concerns and ethical considerations.
			Human-centred values	**Partially fulfilled**	**Although standards under development such as ISO/IEC 42105 provide guidelines for human interaction and oversight, there are no comprehensive guidelines to bring AI systems more in touch with the non-technical users.**
			Privacy protection and security	Fulfilled	Covered by ISO/IEC AWI TS 22443: Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations.
			Reliability and safety	Fulfilled	Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.
			Transparency and explainability	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Accountability	Fulfilled	
			Fairness	Fulfilled	
			Contestability	**Partially fulfilled**	**Although clause 10.2: Nonconformity and corrective**

					action of ISO/IEC 42001: Information technology - Artificial intelligence - Management system provides guidance on taking action and dealing with the consequences in cases of non-conformity, further amendments are required to state the guidelines for individuals seeking correction or remedy if they are negatively affected by decisions made by AI systems.
7.	Model Artificial Intelligence Governance Framework (Second Edition)[7]	Singapore	Risk management	**Partially Fulfilled**	**Standards need to be developed classifying AI systems on the different risk-levels prescribed, and analyzed and assessed for potential threats separately.**
			Data management and protection	Fulfilled	Covered by IS/ISO/IEC 8183: Information technology - Artificial intelligence - Data life cycle framework.
			Explainability, transparency and fairness	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Human-centric AI	**Partially Fulfilled**	**Although standards under development such as ISO/IEC 42105 provide guidelines for human interaction and oversight, there are no comprehensive guidelines to bring AI systems more in touch with the non-technical users.**
8.	National Strategy for Artificial Intelligence [8]	India	Fairness	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254:
			Transparency	Fulfilled	

					Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Privacy	Fulfilled	Covered by ISO/IEC AWI TS 22443: Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations.
			Security	Fulfilled	Covered by ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.
9.	National Guidelines for Artificial Intelligence (AI) Ethics [9]	South Korea	Safeguarding Human Rights	Fulfilled	Covered by ISO/IEC DTS 12791.2: Information technology - Artificial intelligence - Treatment of unwanted bias in classification and regression machine learning tasks and SC42 N1438 NP Outline Guidance Social Ethical concerns Information technology - Artificial intelligence – Guidance for addressing societal concerns and ethical considerations.
			Respect for Diversity	Fulfilled	
			Protection of Privacy	Fulfilled	Covered by ISO/IEC AWI TS 22443: Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations.
			Prevention of Harm	Fulfilled	ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management and ISO/IEC TS 8200: Information technology - Artificial intelligence - Controllability of automated artificial intelligence systems
			Data Management	Fulfilled	Covered by IS/ISO/IEC 8183: Information technology - Artificial intelligence - Data life cycle framework.
			Safety	Fulfilled	ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and

					Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.
			Transparency	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Public Good	Fulfilled	
			Solidarity	Fulfilled	
			Accountability	Fulfilled	
10.	Social Principles of Human-Centric AI Governance Guidelines for Implementation of AI Principles [10]	Japan	Human-Centric	**Partially fulfilled**	**Although standards under development such as ISO/IEC 42105 provide guidelines for human interaction and oversight, there are no comprehensive guidelines to bring AI systems more in touch with the non-technical users.**
			Education/Literacy	**Not fulfilled**	**There are no standards advocating for AI literacy among the various stakeholders**
			Privacy Protection	Fulfilled	Covered by ISO/IEC AWI TS 22443: Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations.
			Ensuring Security	Fulfilled	Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.
			Fair Competition	**Not fulfilled**	**There are no standards advocating for a fair competition in the field of AI**

			Fairness, Accountability, and Transparency	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
11.	Israel's Policy on Artificial Intelligence Regulation and Ethics [11]	Israel	A Risk-Based approach	**Not fulfilled**	**Standards need to be developed classifying AI systems on the different risk-levels prescribed, and analyzed and assessed for potential threats separately.**
			Human-centric AI	**Partially fulfilled**	**Although standards under development such as ISO/IEC 42105 provide guidelines for human interaction and oversight, there are no comprehensive guidelines to bring AI systems more in touch with the non-technical users.**
			Equality and non-discrimination	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Transparency and explainability	Fulfilled	
			Accountability	Fulfilled	
			Reliability, robustness, security and safety	Fulfilled	Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on

					risk management.
12.	Bill 2338/2023 [12]	Brazil	Bias mitigation	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Transparency	Fulfilled	
			Accountability	Fulfilled	
			Traceability of AI systems	Fulfilled	
			Reliability and Robustness of AI		Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.
			Safe and secure AI	Fulfilled	Covered by ISO/IEC AWI TS 22443: Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations.
			Risk-based approach	**Not fulfilled**	**Standards need to be developed classifying AI systems on the different risk-levels prescribed, and analyzed and assessed for potential threats separately.**
13.	Document A/78/L.49 Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development [13]	The UN General Assembly	Safe, secure and trustworthy AI systems	Fulfilled	Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.

			Transparency	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Diversity, equity and inclusion	Fulfilled	
14.	Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law [14]	The European Commission	Risk and Impact Management	**Not fulfilled**	**Standards need to be developed classifying AI systems on the different risk-levels prescribed, and analyzed and assessed for potential threats separately.**
			Human dignity and individual autonomy	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.
			Transparency and oversight	Fulfilled	
			Accountability and responsibility	Fulfilled	
			Equality and non-discrimination	Fulfilled	
			Privacy and personal data protection	Fulfilled	Covered by ISO/IEC AWI TS 22443: Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations.
			Reliability	Fulfilled	Covered by IS/ISO/IEC 24029-2: Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods, ISO/IEC 25059: Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management.

15.	Digital Services Act [15	The European Union	Mitigation of Disinformation	**Not fulfilled**	**There are no comprehensive standards specifically aimed at the authenticity and verification of content generated by AI. Standards related to AI do not directly address these issues**
			Transparency and accountability	Fulfilled	Covered by ISO/IEC 23894: Information technology - Artificial intelligence - Guidance on risk management, ISO/IEC 42001: Information technology - Artificial intelligence - Management system, and ISO/IEC 25059: Software engineering - Systems and software Quality Requirements, Evaluation (SQuaRE) - Quality model for AI systems and ISO/IEC TS 6254: Information technology - Artificial intelligence - Objectives and approaches for explainability of ML models and AI systems.

INFERENCE AND RECOMMENDATIONS

The analysis reveals essential gaps in AI regulations and standards that, if not addressed, will hinder global AI innovation and governance landscape, and the delicate balance between the two. There is a clear need for harmonized standards to ensure ethical practices, risk management, and sector-specific regulations. Without bridging these gaps, the balance between promoting AI advancements and ensuring safe, fair, and responsible AI usage will be disrupted, potentially stalling the progress of AI technologies worldwide.

Based on the analysis, the following recommendations are proposed:

1. **Concrete Guidelines for Risk-based Classification of AI Systems:** Guidelines and frameworks need to be developed to classify AI systems as prohibited or high-risk, ensuring ethical AI practices and protecting user rights.

2. **Standards for Authenticity in AI-Generated Content:** Such standards would make it easier to assess AI-generated content to combat misinformation and enhance consumer protection.

3. **Technical Specifications for AI Model Assessment*:*** Standards establishing technical specifications for assessing AI models' computing power and data center requirements would ensure regulatory compliance and prevent malicious use.

4. **Exclusive Standards for Swarm Intelligence*:*** Exclusive standards on swarm intelligence would enhance large-scale cooperation, perception, and knowledge application, as well as help in regulation of upcoming AI systems based on technology in future.

5. **Standards to ensure fair competition:** Development of standards to ensure fair competition in the field ofAI developments.

6. **Sector-Specific Standards for AI Application Mapping*:*** Development of sector-specific standards would help address the varying levels of risks in different sectors, facilitating effective risk management.

7. **Comprehensive Guidelines for Non-Technical User Accessibility*:*** Such guidelines would make AI systems more accessible to non-technical users, bridging the gap for the 'technology poor' and ensuring equitable AI adoption.

8. **Standards on Technical Documentation and Record keeping:** Development of guidelines focusing on automatic logging of high-risk AI systems.

9. **Standards on mandatory registration of high-risk systems:** Such standards would help enforce many regulations which involve strict regulation of high-risk AI systems.

10. **Standards on Contestability and Redress:** Development of standards aiming to specify strict guidelines for individuals seeking correction or remedy if they are negatively affected by decisions made by AI systems.

INSIGHTS FROM INDUSTRIAL VISIT

SITE: Business Brio
Analytics | AI-ML | NLP | LLM
IT Services and IT Consulting
Kolkata, West Bengal

During my visit to Business Brio on 02.07.2024, I had the opportunity to engage with, Mr. Gautam Banerjee - Founder and Managing Director of Business Brio, and Member, LITD 30 . Our discussion provided invaluable insights into the industry-level functioning of collaborative efforts in data science and AI projects. I gained a deeper understanding of the challenges they face in these domains. Additionally, our conversation shed light on the anticipated complexities AI regulations will introduce upon implementation, emphasizing the practical implications for stakeholders in the field.

I hereby summarize the key elements of our discussion:

1. **Organizations fail to realize the full potential of their Data Science endeavors due to mismanagement and miscommunication between the several teams involved at various levels:** Industry level Data Science and BDA projects face massive challenges due to significant gaps in collaboration and coordination between teams at different levels specializing in separate phases of the process. Mismanagement and miscommunication at these critical junctures hinder seamless integration of insights and implementation of effective strategies, thereby impeding overall project success and outcomes.

2. **Upcoming regulations will make the development of standards extremely crucial:** Upcoming regulations will necessitate the development of standards to ensure compliance, interoperability, and ethical use of technologies like AI. Standardization ensures consistency in practices across industries, facilitates regulatory adherence, promotes innovation within established guidelines, and enhances trust among stakeholders and the public in emerging technologies.

3. **Intellectual Property is everything in the 21st century world:** Intellectual Property (IP) is the goldmine of the 21st century, particularly in the race for global AI leadership. In the AI sector, IP rights safeguard groundbreaking innovations, algorithms, and vast datasets, which are crucial for technological advancement. Nations and corporations that adeptly protect and leverage their IP assets can dominate AI development, drawing significant investments and stimulating further innovation. This competitive advantage not only drives rapid technological progress but also

solidifies a formidable global standing.

4. **<u>Eventually the world should (and will) move on towards the development of sector-specific standards</u>:** Recognizing the diverse needs of different industries is essential. This will provide customized regulations and guidelines, fostering innovation and addressing unique challenges and help tackle risks of varying degrees in each sector, ensuring safety, fairness, and compliance. This evolution towards specialized standards is essential for sustainable growth and technological advancement, enabling industries to thrive within a well-regulated framework.

CONCLUSION

Artificial Intelligence (AI) has become indispensable in our modern lives, permeating industries from healthcare to finance, and transforming how we interact with technology. As AI continues to evolve, effective technical and administrative decision-making processes are crucial to maximize its benefits while mitigating risks. The study underscores the critical need for cohesive global AI regulations and standards to foster innovation responsibly. Bridging gaps between existing regulations and emerging technologies is imperative to ensure ethical AI practices, protect user rights, and maintain a competitive edge in the global AI landscape. Moving forward, harmonized standards will be essential to guide AI development, promote interoperability, and uphold societal trust, thereby paving the way for a sustainable and inclusive AI-driven future.

It is ultimately for the people of this world to decide - whether Artificial Intelligence will be a boon or a bane for us.

BIBLIOGRAPHY

1. https://artificialintelligenceact.eu/wp-content/uploads/2021/08/The-AI-Act.pdf
2. https://assets.publishing.service.gov.uk/media/64cb71a547915a00142a91c4/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf
3. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
4. https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document
5. https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/
6. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles
7. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf
8. https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf
9. https://ai.kisdi.re.kr/eng/main/contents.do?menuNo=500011
10. https://www8.cao.go.jp/cstp/english/humancentricai.pdf
11. https://www.gov.il/BlobFolder/policy/ai_2023/en/Israels%20AI%20Policy%202023.pdf
12. https://clairk.digitalpolicyalert.org/documents/brazil-bill-on-the-use-of-artificial-intelligence-2338-2023-original-language/raw
13. https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf
14. https://rm.coe.int/1680afae3c
15. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065