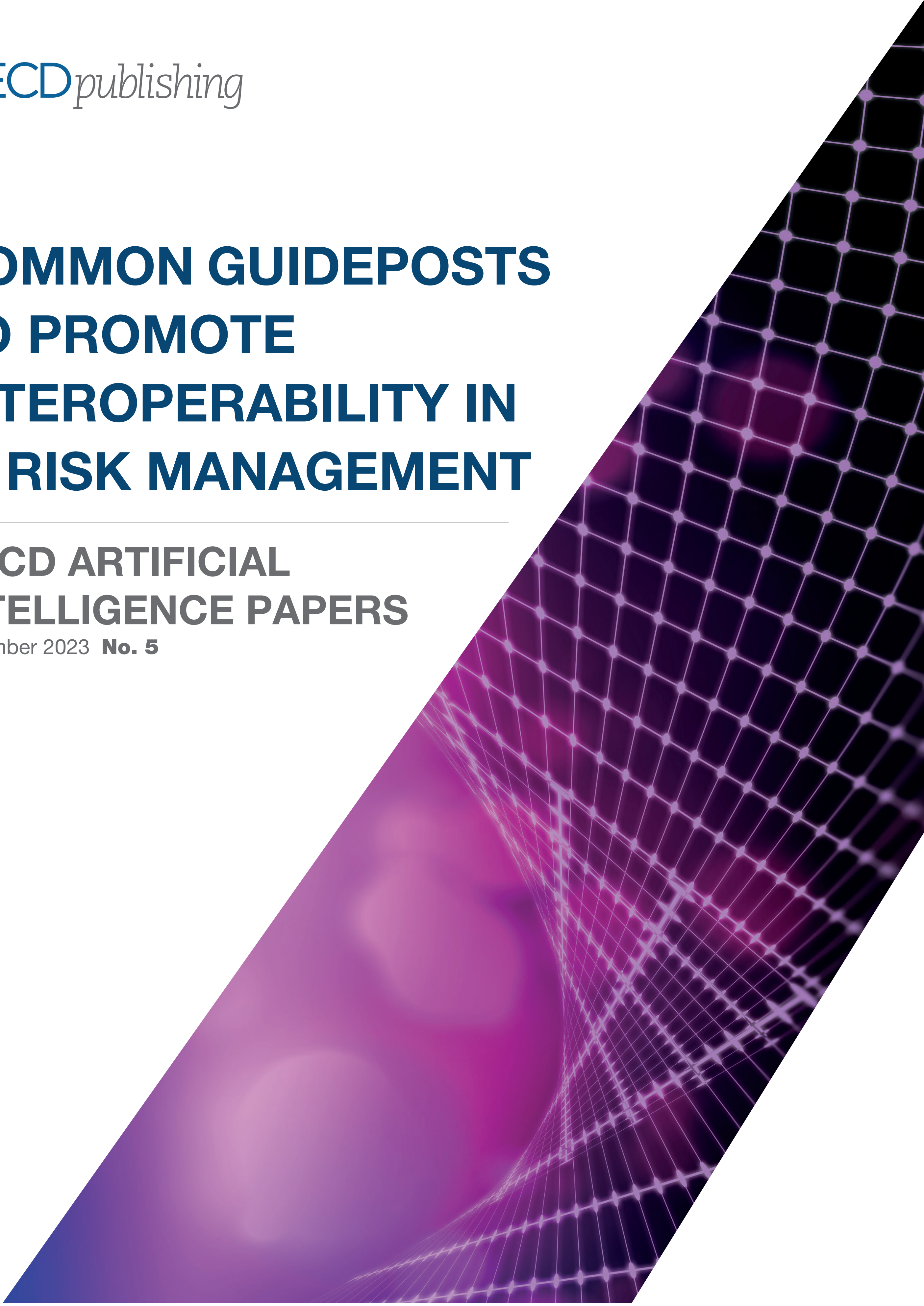
OECDpublishing

COMMON GUIDEPOSTS TO PROMOTE INTEROPERABILITY IN AI RISK MANAGEMENT

OECD ARTIFICIAL INTELLIGENCE PAPERS

November 2023 **No. 5**



Federal Ministry
of Labour and Social Affairs

OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

The report provides a high-level overview and analysis of commonalities and differences of leading risk management frameworks for AI. The report aims at developing a common understanding of the AI risk and accountability landscape, with the ultimate objective of operationalising the OECD AI Principles and OECD instruments on responsible business conduct (RBC) in the AI sector.

This report was discussed and reviewed by the OECD Working Party on Artificial Intelligence (AIGO) and OECD.AI Expert Group on Risk & Accountability in April and June 2023.

This report contributes to the OECD's Artificial Intelligence in Work, Innovation, Productivity and Skills (AI-WIPS) programme, which provides policy makers with new evidence and analysis to keep abreast of the fast-evolving changes in AI capabilities and diffusion and their implications for the world of work. For more information, please visit www.oecd.ai/wips. AI-WIPS is supported by the German Federal Ministry of Labour and Social Affairs (BMAS) and will complement the work of the German AI Observatory in the Ministry's Policy Lab Digital, Work & Society. For more information, visit https://oecd.ai/work-innovationproductivity-skills and https://denkfabrik-bmas.de/.

This report was written by Karine Perset, Luis Aranda, and Rashad Abelson under the supervision of Audrey Plonk, Head of the OECD Digital Economy Policy Division. The report also benefitted from the inputs of delegates for the OECD Working Party on Artificial Intelligence (AIGO), including the Civil Society Information Society Advisory Council (CSISAC) and Business at the OECD (BIAC). Orsolya Dobe and Shellie Phillips provided editorial support.

This report was prepared for publication by the OECD Secretariat in consultation with the delegates of the Working Party on Artificial Intelligence Governance (AIGO). The report was approved and declassified by written procedure by the Committee on Digital Economy Policy on 12/10/2023.

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CDEP/AIGO(2022)10/FINAL

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Acknowledgements

This report is based on the work of the OECD.AI Expert Group on Risk & Accountability. It was prepared under the aegis of the OECD Working Party on AI Governance (AIGO). The Expert Group is co-chaired by Nozha Boujemaa (Decathlon); Andrea Renda (Centre for European Policy Studies); and Sebastian Hallensleben (CEN-CENELEC). Luis Aranda, Rashad Abelson and Karine Perset (OECD Digital Economy Policy Division), led the report development and drafting.

Around 100 experts participated in regular virtual and in-person meetings between February 2023 and August 2023. Many experts provided invaluable feedback and suggestions. The report benefitted from initial drafting and research from Olivia Erdelyi (University of Canterbury). The report also benefitted significantly from the contributions of national delegations to the OECD Working Party on AI Governance.

The Secretariat would also like to thank stakeholder groups at the OECD for their input, including Pam Dixon (Civil Society Information Society Advisory - CSISAC); Christina Colclough (Trade Union Advisory Committee – TUAC); and Nicole Primmer and Maylis Berviller (Business at OECD – BIAC).

Finally, the authors thank Orsolya Dobe and Shellie Phillips for editorial support. The overall quality of this report benefitted significantly from their engagement.

Table of contents

Tables

Figures

Boxes

Abstract

The OECD AI Principles call for AI actors to be accountable for the proper functioning of their AI systems in accordance with their role, context, and ability to act. Likewise, the OECD Guidelines for Multinational Enterprises sets out the government-backed expectation that all businesses avoid and address negative impacts of their operations, while contributing to sustainable development in the countries where they operate. To develop 'trustworthy' and 'responsible' AI systems, there is a need to identify and treat AI risks. As governments, experts and other stakeholders increasingly call for the development of accountability mechanisms, namely through risk management frameworks, interoperability between burgeoning frameworks would be desirable to help increase efficiencies and reduce enforcement and compliance costs. This report provides a high-level analysis of the commonalities and differences of leading AI risk management frameworks currently under development. This report demonstrates that while the order of the risk management steps, the target audience, scope and specific terminology sometimes differ, all the risk management frameworks analysed follow a similar and sometimes functionally equivalent risk management process.

Résumé

Les Principes sur l’IA de l’OCDE appellent les acteurs de l’IA à être responsables du bon fonctionnement de leurs systèmes d’IA conformément à leur rôle, du contexte et leur capacité d’action. De même, les Principes directeurs de l'OCDE à l'intention des entreprises multinationales exposent l'attente soutenue par les gouvernements selon laquelle toutes les entreprises évitent et traitent les impacts négatifs de leurs activités, tout en contribuant au développement durable dans les pays où elles sont implantées. Pour développer des systèmes d’IA « dignes de confiance » et « responsables », il est nécessaire d’identifier et de gérer les risques liés à l’IA. Alors que les gouvernements, les experts et d’autres parties prenantes appellent de plus en plus à l’élaboration de mécanismes de responsabilisation, notamment par le biais de cadres de gestion des risques, l’interopérabilité entre les cadres et les standards en plein essor serait souhaitable pour contribuer à accroître l’efficacité et à réduire les coûts d’application et de conformité. Ce rapport fournit une analyse de haut niveau des points communs et des différences entre les principaux cadres de gestion des risques liés à l’IA actuellement en cours d’élaboration. Ce rapport démontre que même si l'ordre des étapes de gestion des risques, le public cible, la portée et la terminologie spécifique diffèrent parfois, tous les cadres de gestion des risques liés à l'IA analysés suivent un processus de gestion des risques similaire et parfois équivalent sur le plan fonctionnel.

Übersicht

Die KI-Grundsätze der OECD fordern, dass KI-Akteur:innen entsprechend ihrer Rolle, ihrem Kontext und ihrer Handlungsfähigkeit für das ordnungsgemäße Funktionieren ihrer KI-Systeme verantwortlich sind. Ebenso legen die OECD-Grundsätze für Multinationale Unternehmen die von der Regierung unterstützte Erwartung dar, dass alle Unternehmen negative Auswirkungen ihrer Geschäftstätigkeit vermeiden und angehen und gleichzeitig zu einer nachhaltigen Entwicklung in den Ländern beitragen, in denen sie tätig sind. Um „vertrauenswürdige“ und „verantwortungsvolle“ KI-Systeme zu entwickeln, müssen KI-Risiken identifiziert und gemanagt werden.Da Regierungen, Expert:innen und andere Interessengruppen zunehmend die Entwicklung von Rechenschaftsmechanismen fordern, insbesondere durch Rahmenwerke für das Risikomanagement, wäre eine Interoperabilität zwischen aufkeimenden Rahmenwerken und Standards wünschenswert, um die Effizienz zu steigern und die Durchsetzungs- und Compliance-Kosten zu senken. Dieser Bericht bietet eine umfassende Analyse der Gemeinsamkeiten und Unterschiede der führenden KI-Risikomanagement-Rahmenwerke, die derzeit entwickelt werden. Dieser Bericht zeigt, dass die Reihenfolge der Risikomanagementschritte, die Zielgruppe, der Umfang und die spezifische Terminologie zwar manchmal unterschiedlich sind, alle analysierten KI-Risikomanagement-Rahmenwerke jedoch einem ähnlichen und manchmal funktional gleichwertigen Risikomanagementprozess folgen.

Background and objectives

Through the OECD.AI Network of experts work stream on AI Risk & Accountability (see Annex A and Annex B), the OECD is engaging with partner organisations, policy makers and experts, to identify common guideposts to assess AI risk and impact for trustworthy AI. The goal is to help implement effective, accountable and trustworthy AI systems by promoting global consistency.

The work stream will:

1. Map existing and developing core standards, frameworks and guidelines for AI risk management to the top-level interoperability framework developed in the report "*Advancing Accountability in AI: Governing and Managing risks through the lifecycle for trustworthy AI*" (OECD, 2023[1]). These include frameworks from major actors like the International Organization for Standardization (ISO)[1], the Institute of Electrical and Electronics Engineers (IEEE), the National Institute of Standards and Technology (NIST), the European Committee for Electrotechnical Standardization (CEN-CENELEC), the OECD, the Government of Canada, the European Commission and the Council of Europe.
2. One level down, take stock of commonalities and differences in concepts and terminology between initiatives and conduct a gap analysis, proposing possible terminology if appropriate, and map the relevant actors in the AI value chain and relevant high-priority risks that frameworks seek to address.
3. Translate analysis into good practice to inform development of due diligence guidance for responsible business conduct (RBC) in AI.
4. Research and analyse the alignment of certification schemes with OECD RBC and AI standards.
5. Develop an interactive online tool to help organisations and stakeholders compare frameworks (see 1 and 2 above) and navigate existing methods, tools and good practices for identifying, assessing, treating and governing AI risks.[2]

This report represents step (1) "mapping core standards, frameworks and guidelines for AI risk management".

Executive summary

Comparing AI risk management standards and frameworks is a first step towards greater consistency and interoperability.

This report provides a high-level overview and analysis of commonalities and differences of leading relevant risk management frameworks for AI. It does so by mapping the Interoperability Framework from the report "*Advancing Accountability in AI: Governing and managing risks through the lifecycle for trustworthy AI*" (OECD, 2023[1]) to existing and draft risk management standards to identify where they are functionally equivalent and where they differ. This report is part of a larger project that seeks to develop a common understanding of the AI accountability ecosystem, in view of fostering coherence and alignment towards common government-reviewed risk management guideposts on trustworthy AI.

Key risk management frameworks are generally aligned with four top-level steps: 'DEFINE', 'ASSESS', and 'TREAT' risks, and 'GOVERN' risk management processes.

In evaluating these risk management frameworks, the report finds general alignment between the Interoperability Framework at the top-level and the different frameworks. While the order of operations, target audience, risk scope, segment of the AI system lifecycle and specific terminology used may differ, all the frameworks generally seek to achieve the same outcomes (responsible, ethical, trustworthy AI) through roughly the same risk management process that includes the following steps:

- 'DEFINE' scope, context and criteria, including the relevant AI principles and risks, stakeholders and actors for each phase of the AI system lifecycle and the lifecycle itself.
- 'ASSESS' the risks to trustworthy AI by identifying and analysing issues at individual, aggregate and societal levels and evaluating the likelihood and level of harm.
- 'TREAT' risks to cease, prevent, or mitigate adverse impacts commensurate with the likelihood and severity of each.
- 'GOVERN' the risk management process by '*embedding*' and cultivating a culture of risk management in organisations; '*monitoring and reviewing*' the process in an ongoing manner; and '*documenting*', '*communicating*' and '*consulting*' on the process and its outcomes, as well as a clear definition or assignment of roles and responsibilities of different AI actors and establishing a line of accountability.

One overarching distinguishing feature among the standards is the primary target of their implementation. The OECD DDG and ISO standards are primarily aimed at board-level or organisational-level changes to enable risk management. While the other standards also offer board-level recommendations, their implementation are primarily at the technical-level (e.g., identifying and addressing risks in AI system design and along the AI system lifecycle).

Most of the differences between frameworks relate to the 'GOVERN' function.

Frameworks vary in nature and scope, which results in different approaches to governing the risk management system. While in some frameworks, governance activities are explicitly included under a 'GOVERN' function, in others they are absent or distributed throughout the risk management process.

While terminology and sequence differ somewhat, the high-level steps of the OECD Due Diligence Guidance and ISO 31000:2018 Risk Management - Guidelines (ISO 31000) map closely to the Interoperability Framework. ISO/IEC 23894 supplements ISO 31000 with AI-specific guidance, but operates within the same top-level framework.

There is a one-to-one mapping of the NIST AI Risk Management Framework's high-level categories with those of the Interoperability Framework. While form may differ between both frameworks' GOVERN function (e.g., in NIST AI RMF, the sub-elements of GOVERN such as 'document', 'monitor', 'consult', and 'communicate' are integrated throughout the different steps in a less structured manner), content is largely equivalent.

In the proposed EU AI Act (EU AIA) and Canada AI and Data Act (AIDA) providers of high-risk AI systems are required to identify, analyse and mitigate risks. Yet, some GOVERN risk management measures from the Interoperability Framework – like consulting with stakeholders and embedding risk management into organisational culture – seem absent from the proposed legislative acts so far. A unique feature of both of proposed acts is that the regulator takes on a number of top-level risk definition and assessment tasks by first determining what is considered to be a high-risk system. This provides a de-facto risk prioritisation mechanism for companies.

The Council of Europe's draft Human Rights, Democracy and the Rule of Law Risk and Impact Assessment (HUDERIA) is partly aligned with the Interoperability Framework, but elements from the Interoperability Framework relating to GOVERN do not seem to be present in HUDERIA. This includes communicating publicly whether an AI system conforms to regulatory, governance and ethical standards after assessing and treating risks, and involving leadership to embed the risk management process across the organisational structure.

The IEEE 7000-21 targets integration of value-based considerations and stakeholder views into product or service design. As such, its scope is narrower than other risk management frameworks and mapping the top-level steps directly with those of the Interoperability Framework is challenging.

ISO/IEC Guide 51 is aimed at informing the development of other standards seeking to integrate product safety requirements in their risk management frameworks. It focuses on risk identification, assessment and reduction and as such is broadly consistent with DEFINE, ASSESS and TREAT in the Interoperability Framework. Some sub-elements of GOVERN, such as embedding risk management policies and consulting with stakeholders, are not included in ISO/IEC Guide 51.

Synthèse

La comparaison des standards et des cadres de gestion des risques liés à l'IA est une première étape vers une plus grande cohérence et interopérabilité.

Ce rapport fournit un aperçu et une analyse de haut niveau des points communs et des différences entre les principaux cadres de gestion des risques pertinents pour l'IA. Pour ce faire, il établit une correspondance entre le cadre d'interopérabilité du rapport « Advancing Accountability in AI : Governing and Management Risks Through the Lifecycle for Trustworthy AI » (OCDE, 2023[1]) et les standards existants – ou à l'état de projet – de gestion des risques afin d'identifier les points fonctionnellement équivalents et ceux qui diffèrent. Ce rapport fait partie d'un projet plus vaste visant à développer une compréhension commune de l'écosystème de responsabilité en matière d'IA, en vue de favoriser la cohérence et l'alignement sur des lignes directrices communes de gestion des risques sur une IA digne de confiance, examinées par les pouvoirs publics.

Les principaux cadres de gestion des risques sont généralement alignés sur 4 étapes de niveau supérieur : « DÉFINIR », « ÉVALUER », « TRAITER » les risques et « GOUVERNER » les processus de gestion des risques.

En évaluant ces cadres de gestion des risques, le rapport constate un alignement général entre le Cadre d'Interopérabilité au plus haut niveau et les différents cadres. Bien que l'ordre des opérations, le public cible, l'étendue des risques, la partie considérée du cycle de vie du système d'IA et la terminologie spécifique utilisée puissent différer, tous les cadres cherchent généralement à atteindre les mêmes résultats (IA responsable, éthique et digne de confiance) par le biais d'un processus de gestion des risques à peu près identique, qui comprend les étapes suivantes :

- « DÉFINIR » la portée, le contexte et les critères, y compris les principes et les risques de l'IA, les parties prenantes et les acteurs pour chaque phase du cycle de vie du système d'IA et le cycle de vie lui-même.
- « ÉVALUER » les risques liés à une IA digne de confiance en identifiant et en analysant les problèmes aux niveaux individuel, global et sociétal et en évaluant la probabilité et le niveau de préjudice.
- « TRAITER » les risques pour cesser, prévenir ou atténuer les impacts négatifs en fonction de la probabilité et de la gravité de chacun.
- « GOUVERNER » le processus de gestion des risques en « intégrant » et en cultivant une culture de gestion des risques dans les organisations ; « surveiller et examiner » le processus de manière continue ; et « documenter », « communiquer » et « consulter » sur le processus et ses résultats, ainsi qu'en définissant ou en attribuant clairement les rôles et les responsabilités des différents acteurs de l'IA et en établissant une ligne de responsabilité.

L’une des principales caractéristiques distinctives des standards est l’objectif principal de leur mise en œuvre. Les standards DDG et ISO de l'OCDE visent principalement des changements au niveau du conseil d'administration ou au niveau plus large de l'organisation pour permettre la gestion des risques. Bien que les autres standards proposent également des recommandations au niveau du conseil d'administration, leur mise en œuvre se situe principalement au niveau technique (e.g., identifier et gérer les risques dans la conception des systèmes d'IA et tout au long du cycle de vie des systèmes d'IA).

La plupart des différences entre les cadres concernent la fonction « GOUVERNER ».

Les cadres varient en nature et en portée, ce qui se traduit par des approches différentes pour gouverner le système de gestion des risques. Alors que dans certains cadres, les activités de gouvernance sont explicitement incluses dans une fonction « GOUVERNER », dans d’autres, elles sont absentes ou réparties tout au long du processus de gestion des risques.

Bien que la terminologie et l’ordre diffèrent quelque peu, les étapes de haut niveau du « Guide OCDE sur le devoir de diligence » et du standard « ISO 31000:2018 Gestion des risques - Lignes directrices (ISO 31000) » correspondent étroitement au cadre d'interopérabilité. Le standard ISO/IEC 23894 complète le standard ISO 31000 avec des lignes directrices spécifiques à l'IA, mais fonctionne dans le même cadre de niveau supérieur.

Les catégories de haut niveau du cadre de gestion des risques liés à l'IA du NIST correspondent à celles du Cadre d'Interopérabilité. Bien que la forme puisse différer entre la fonction GOUVERNER des deux cadres (par exemple dans « NIST AI RMF », les sous-éléments de GOUVERNER tels que « documenter », « surveiller », « consulter » et « communiquer » sont intégrés tout au long des différentes étapes de manière de manière moins structurée), le contenu est largement équivalent.

Dans l’ « EU AI Act » (EU AIA) et dans le « Canada AI and Data Act » (AIDA), les fournisseurs de systèmes d’IA à haut risque sont tenus d’identifier, d’analyser et d’atténuer les risques. Pourtant, certaines mesures de gestion des risques du cadre d’interopérabilité GOUVERNER – comme la consultation des parties prenantes et l’intégration de la gestion des risques dans la culture organisationnelle – semblent jusqu’à présent absentes des actes législatifs proposés. Une caractéristique unique ces deux lois proposées est que l'autorité de régulation se charge d’un certain nombre de tâches de définition et d'évaluation des risques de haut niveau en déterminant d'abord ce qui est considéré comme un système à haut risque. Cela fournit de facto un mécanisme de priorisation des risques pour les entreprises.

Le projet d’évaluation des risques et de l’impact en matière des droits de l’homme, de la démocratie et de l’État de droit du Conseil de l’Europe (HUDERIA) est en partie aligné sur le Cadre d’Interopérabilité, mais les éléments du cadre d’interopérabilité relatifs à GOUVERNER ne semblent pas être présents dans l’HUDERIA. Il s'agit notamment d'indiquer publiquement si un système d’IA est conforme aux normes réglementaires, de gouvernance et d’éthiques après avoir évalué et traité les risques, et d’impliquer les dirigeants pour intégrer le processus de gestion des risques dans l’ensemble de la structure organisationnelle.

L'IEEE 7000-21 vise à intégrer des considérations basées sur la valeur et des points de vue des parties prenantes dans la conception de produits ou de services. Sa portée est donc plus étroite que celle des autres cadres de gestion des risques et il est difficile de faire correspondre les étapes de haut niveau directement avec celles du Cadre d'Interopérabilité.

Le Guide ISO/CEI 51 vise à informer l'élaboration d'autres normes cherchant à intégrer les exigences de sécurité des produits dans leurs cadres de gestion des risques. Il se concentre sur l’identification, l’évaluation et la réduction des risques et, à ce titre, est globalement conforme aux principes DEFINIR, ÉVALUER et TRAITER du Cadre d’Interopérabilité. Certains sous-éléments de GOUVERNER, tels que l'intégration de politiques de gestion des risques et la consultation des parties prenantes, ne sont pas inclus dans le Guide ISO/IEC 51.

Zusammenfassung

Der Vergleich von Standards und Rahmenwerken für das KI-Risikomanagement ist ein erster Schritt zu mehr Konsistenz und Interoperabilität.

Dieser Bericht bietet einen umfassenden Überblick und eine Analyse der Gemeinsamkeiten und Unterschiede der führenden relevanten Risikomanagement-Rahmenwerke für KI. Dies geschieht durch die Zuordnung des Interoperabilitätsrahmens aus dem Bericht „*Advancing Accountability in AI: Governing and manageing hazards through the lifecycle for trustworthy AI*“ (OECD, 2023[1]) zu bestehenden und entworfenen Risikomanagementstandards, um festzustellen, wo sie funktional gleichwertig sind und wo sie sich unterscheiden. Dieser Bericht ist Teil eines größeren Projekts, das darauf abzielt, ein gemeinsames Verständnis des Ökosystems der KI-Rechenschaftspflicht zu entwickeln, um die Kohärenz und Ausrichtung auf gemeinsame, von der Regierung überprüfte Risikomanagement-Leitfäden für vertrauenswürdige KI zu fördern.

Wichtige Rahmenwerke für das Risikomanagement sind im Allgemeinen auf vier Schritte der obersten Ebene ausgerichtet: Risiken „DEFINIEREN“, „BEWERTEN“ und „BEHANDELN“ sowie Risikomanagementprozesse „LENKEN“.

Bei der Bewertung dieser Risikomanagement-Frameworks stellt der Bericht eine allgemeine Übereinstimmung zwischen dem Interoperabilitäts-Framework auf der obersten Ebene und den verschiedenen Frameworks fest. Auch wenn die Reihenfolge der Abläufe, die Zielgruppe, der Risikoumfang, der Abschnitt des Lebenszyklus des KI-Systems und die verwendete spezifische Terminologie unterschiedlich sein können, zielen alle Rahmenwerke im Allgemeinen darauf ab, die gleichen Ergebnisse (verantwortungsvolle, ethische, vertrauenswürdige KI) durch ungefähr denselben Risikomanagementprozess zu erzielen. Dazu gehören folgende Schritte:

- „DEFINIEREN“ Sie Umfang, Kontext und Kriterien, einschließlich der relevanten KI-Prinzipien und -Risiken, Stakeholder und Akteur:innen für jede Phase des KI-Systemlebenszyklus und den Lebenszyklus selbst.

- „BEWERTEN“ Sie die Risiken für eine vertrauenswürdige KI, indem Sie Probleme auf individueller, aggregierter und gesellschaftlicher Ebene identifizieren und analysieren, und die Wahrscheinlichkeit und das Ausmaß des Schadens bewerten.

- „BEHANDELN“ Sie Risiken, um nachteilige Auswirkungen entsprechend ihrer Wahrscheinlichkeit und Schwere zu stoppen, zu verhindern oder zu mildern.

- „LENKEN“ Sie den Risikomanagementprozess, indem Sie eine Kultur des Risikomanagements in Organisationen „einbetten“ und kultivieren. den Prozess fortlaufend „überwachen und überprüfen“; und „Dokumentieren“, „Kommunizieren“ und „Beraten“ des Prozesses und seiner Ergebnisse sowie eine klare Definition oder Zuweisung von Rollen und Verantwortlichkeiten verschiedener KI-Akteur:innen und die Festlegung einer „Verantwortlichkeitslinie“.

Ein übergeordnetes Unterscheidungsmerkmal der Standards ist das primäre Ziel ihrer Umsetzung. Die OECD-DDG- und ISO-Standards zielen in erster Linie auf Änderungen auf Vorstands- oder Organisationsebene ab, um ein Risikomanagement zu ermöglichen. Während die anderen Standards auch Empfehlungen auf Vorstandsebene beinhalten, erfolgt ihre Umsetzung hauptsächlich auf technischer Ebene (z. B. Identifizierung und Bewältigung von Risiken beim KI-Systemdesign und entlang des KI-Systemlebenszyklus).

Die meisten Unterschiede zwischen Frameworks beziehen sich auf die Funktion „LENKEN“.

Die Rahmenwerke variieren in Art und Umfang, was zu unterschiedlichen Ansätzen zur Steuerung des Risikomanagementsystems führt. Während in einigen Rahmenwerken Governance-Aktivitäten ausdrücklich in einer „LENKUNGS“-Funktion enthalten sind, fehlen sie in anderen oder sind auf den gesamten Risikomanagementprozess verteilt.

Obwohl sich Terminologie und Reihenfolge etwas unterscheiden, sind die übergeordneten Schritte der OECD-Leitlinien zur Sorgfaltspflicht und der Risikomanagement-Leitlinien ISO 31000:2018 (ISO 31000) eng mit dem Interoperabilitätsrahmen verknüpft. ISO/IEC 23894 ergänzt ISO 31000 durch KI-spezifische Leitlinien, operiert jedoch innerhalb desselben Rahmenwerks der obersten Ebene.

Es gibt eine Eins-zu-eins-Zuordnung der übergeordneten Kategorien des NIST AI Risk Management Framework zu denen des Interoperability Framework. Während sich die Form der LENKUNGS-Funktion beider Frameworks unterscheiden kann (z. B. in NIST AI RMF, die Unterelemente von LENKUNG wie „Dokumentieren“, „Überwachung“, „Beratung“ und „Kommunikation“ sind in den verschiedenen Schritten integriert strukturiert) sind die Inhalte weitgehend gleichwertig.

Im vorgeschlagenen *EU AI Act (EU AIA)* und *Canada AI and Data Act (AIDA)* sind Anbieter:innen von Hochrisiko-KI-Systemen verpflichtet, Risiken zu identifizieren, zu analysieren und zu mindern. Dennoch scheinen einige LENKUNGS-Risikomanagementmaßnahmen aus dem Interoperabilitätsrahmen – wie die Konsultation von Interessengruppen und die Einbettung des Risikomanagements in die Organisationskultur – in den vorgeschlagenen Gesetzen bisher nicht enthalten zu sein. Ein einzigartiges Merkmal beider vorgeschlagenen Gesetze besteht darin, dass die Regulierungsbehörde eine Reihe von Aufgaben zur Risikodefinition und -bewertung auf höchster Ebene übernimmt, indem sie zunächst festlegt, was als Hochrisikosystem gilt. Dies bietet Unternehmen einen faktischen Risikopriorisierungsmechanismus.

Der Entwurf des Europarates zur Risiko- und Folgenabschätzung für Menschenrechte, Demokratie und Rechtsstaatlichkeit (HUDERIA) ist teilweise auf den Interoperabilitätsrahmen abgestimmt, Elemente des Interoperabilitätsrahmens in Bezug auf LENKUNG scheinen jedoch in HUDERIA nicht vorhanden zu sein. Dazu gehört die öffentliche Kommunikation darüber, ob ein KI-System nach der Bewertung und Behandlung von Risiken den regulatorischen, Governance- und ethischen Standards entspricht, und die Einbeziehung der Führung, um den Risikomanagementprozess in der gesamten Organisationsstruktur zu verankern.

Die IEEE 7000-21 zielt auf die Integration wertbasierter Überlegungen und Stakeholder:innen-Ansichten in das Produkt- oder Servicedesign ab. Daher ist sein Anwendungsbereich enger als bei anderen Risikomanagement-Frameworks. Außerdem stellt die direkte Zuordnung der obersten Schritte zu denen des Interoperabilitäts-Frameworks eine Herausforderung dar.

Der ISO/IEC-Leitfaden 51 soll als Grundlage für die Entwicklung anderer Standards dienen, die darauf abzielen, Produktsicherheitsanforderungen in ihre Risikomanagementrahmen zu integrieren. Er konzentriert sich auf die Identifizierung, Bewertung und Reduzierung von Risiken und steht daher weitgehend im Einklang mit DEFINIEREN, BEWERTEN und BEHANDELN im Interoperabilitätsrahmen. Einige Unterelemente von LENKEN, wie die Einbettung von Risikomanagementrichtlinien und die Konsultation von Interessengruppen, sind im ISO/IEC Guide 51 nicht enthalten.

1 Mapping top-level interoperability between frameworks

While AI provides tremendous benefits, it also presents real risks like bias and discrimination, the polarisation of opinions, privacy infringement and widespread surveillance in some countries. Some of these risks are already materialising into harms to people and society. To develop ‘trustworthy’, ‘responsible’ or ‘ethical’ AI systems, there is a need to assess impacts and manage AI risks. Over the past few years, there has been global convergence towards using – voluntary or mandatory – risk-based approaches and impact assessments to help govern AI. Demand is growing in the public and private sectors for tools and processes to help document AI system decisions and facilitate accountability throughout the AI system lifecycle – from planning and design, to data collection and processing, to model building and validation, to deployment, operation and monitoring.

At the same time, interoperability between burgeoning frameworks and standards is desirable, ideally ahead of their implementation in mandatory and voluntary AI risk assessment and management standards. A proliferation of different frameworks and standards that are not interoperable could make implementation of trustworthy AI more complex and costly in practice and therefore less effective and less enforceable. Facilitating such interoperability calls for co-operation and coordination between domestic and international state and non-state actors developing standards and frameworks on AI risk management; AI design (e.g., trustworthiness by design); and AI impact, conformity and risk assessments.

One of the ten OECD AI Principles refers to the accountability that AI actors bear for the proper functioning of the AI systems they develop and use (Box 1. What is trustworthy AI?). To remain accountable, AI actors need to govern and manage risks[3] throughout their AI systems’ lifecycle. The OECD report “*Advancing Accountability in AI: Governing and managing risks through the lifecycle for trustworthy AI*” (OECD, 2023[1]) illustrates how risk management approaches can provide a systematic way to do so. The ‘High-Level AI Risk Management Interoperability Framework’ (Interoperability Framework) draws from existing standards to identify four essential risk management steps for AI (Figure 1. High-level AI risk management interoperability framework

- **DEFINE** scope, context and criteria, including the relevant AI principles, stakeholders and actors for each phase of the AI system lifecycle and for the lifecycle itself.
- **ASSESS** the risks to trustworthy AI by identifying and analysing issues at individual, aggregate and societal levels and evaluating the likelihood and level of harm (e.g., small risks can add up to larger risk).
- **TREAT** risks to cease, prevent, or mitigate adverse impacts, commensurate with the likelihood and severity of each.
- **GOVERN** the risk management process by embedding and cultivating a culture of risk management in organisations; monitoring and reviewing the process in an ongoing manner; and documenting, communicating and consulting on the process and its outcomes.

Providing accountability for trustworthy AI requires that actors leverage processes, indicators, standards, certification schemes, auditing and other mechanisms to follow these steps at each phase of the AI system lifecycle, which encompasses the following phases: (1) plan and design; (2) collect and process data; (3)

build and use the model; (4) verify and validate the model; (5) deploy (including ‘putting into service’ and ‘placing the AI system on the market’)[4]; and (6) operate and monitor the system, which may include retiring an AI system from operation (OECD, 2019[2]; OECD, 2022[3]). This should be an iterative process where the findings and outputs of one risk management stage feed into the others.

To develop a clear understanding of the accountability ecosystem and to put the Interoperability Framework into practice, this report maps relevant standards to the Interoperability Framework at the top structural levels (Table 1. High-level mapping of selected risk management frameworks for AI to the Interoperability Framework). Each of the following subsections provides an overview of the covered standards, including their objectives, scope of coverage and top-level risk management processes, followed by a brief analysis of the commonalities and gaps between the standard and the Interoperability Framework. The analysis in this report will focus only on process. Differences in terminology will be the focus of future reports.

While numerous standards exist or are currently being developed[5], the standards assessed in this report, given their relevance to AI risk management, are:

- OECD Due Diligence Guidance for Responsible Business Conduct (OECD DDG)
- ISO 31000:2018 Risk Management - Guidelines (ISO 31000) + ISO/IEC 23894:2023
- United States National Institute of Standards and Technology, AI Risk Management Framework (NIST AI RMF)
- European Commission proposal for a Regulation laying down harmonised rules on AI (EU AIA)
- Government of Canada proposed Artificial Intelligence and Data Act (AIDA)
- Council of Europe Human Rights, Democracy and the Rule of Law Assurance Framework for AI Systems (HUDERIA)
- Institute of Electrical and Electronics Engineers 7000-21 Standard Model Process for Addressing Ethical Concerns during System Design (IEEE 7000-21)
- ISO/IEC Guide 51:2014, Third Edition (ISO/IEC Guide 51)

One overarching distinguishing feature among the standards is the primary target of their implementation. The OECD DDG and ISO standards are primarily aimed at board-level or organisational-level changes to enable risk management. While the other standards also offer board-level recommendations, their implementation are primarily at the technical-level (e.g., identifying and addressing risks in AI system design and along the AI system lifecycle).

Box 1. What is trustworthy AI?

In this report, “trustworthy AI” refers to systems that embody the OECD’s values-based AI Principles:

- **Benefiting people and the planet:** Those who play an active role in the AI system lifecycle (AI actors) and stakeholders, including civil society and affected communities, should engage in creating AI systems that can contribute to inducing inclusive growth, sustainable development and wellbeing.
- **Human-centred values and fairness:** AI actors should respect the rule of law, human rights, and democratic values throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights. To that end, AI actors should implement mechanisms and safeguards that are appropriate to the context and consistent with the state of art.
- **Transparency and explainability:** AI actors, including organisations and individuals that deploy or operate AI, should commit to responsible disclosures to make stakeholders aware of their interactions with AI systems and provide information to foster stakeholders’ understanding of the systems, such that people affected by AI systems can comprehend the outcome and challenge the decision when needed.
- **Robustness, security and safety:** AI systems need to function appropriately while ensuring traceability and AI actors need to apply systematic risk management approaches to mitigate, among others, safety and security risks.
- **Accountability:** AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context and consistent with the state of art.

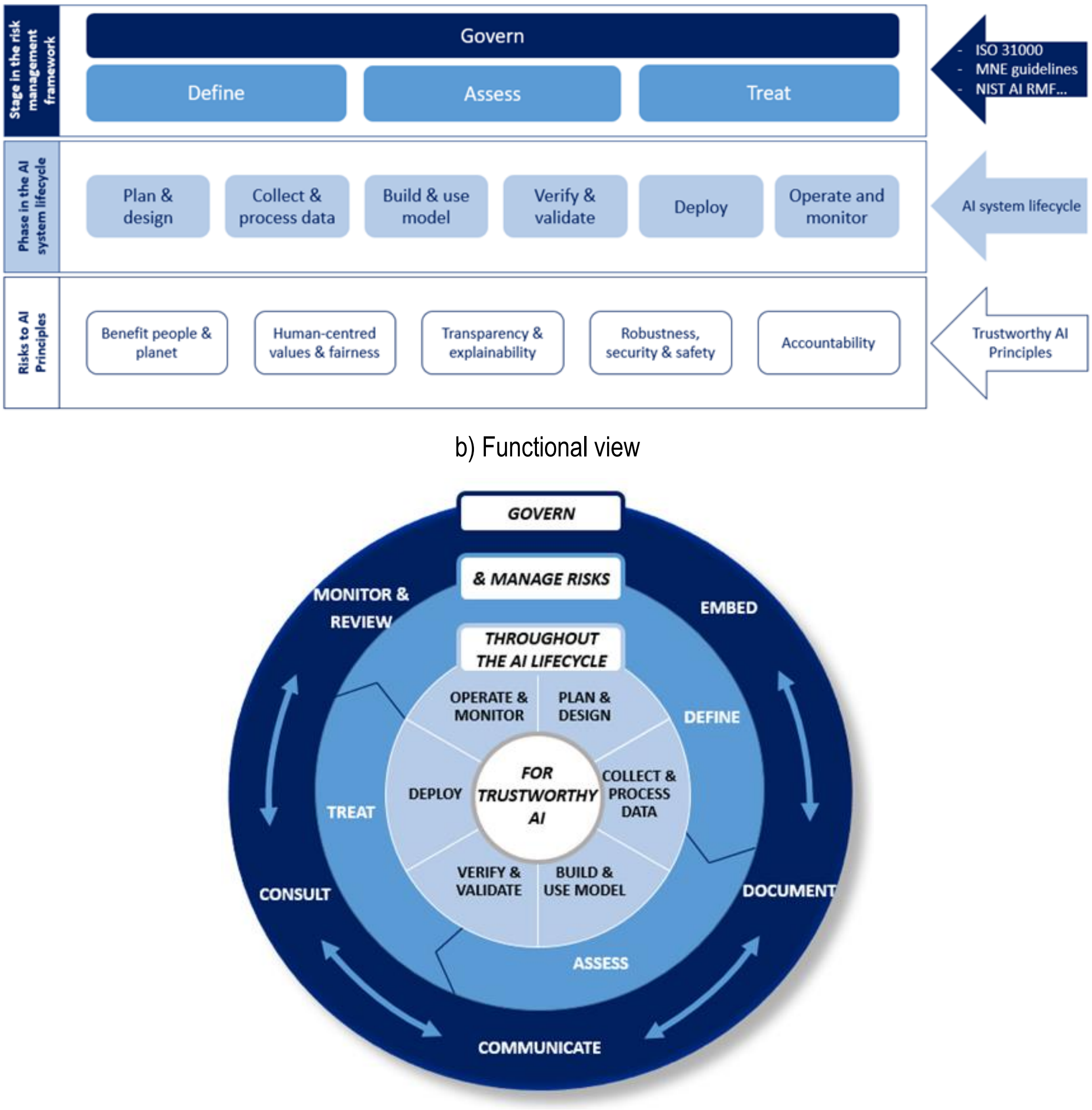
Governments, international organisations, civil society, and companies are increasingly working on developing frameworks, guidance, best practice and other mechanisms to enable and verify the development, deployment and use of trustworthy AI.

Source: (OECD, 2019[4])

Figure 1. High-level AI risk management interoperability framework

Governing and managing risks throughout the lifecycle for trustworthy AI.

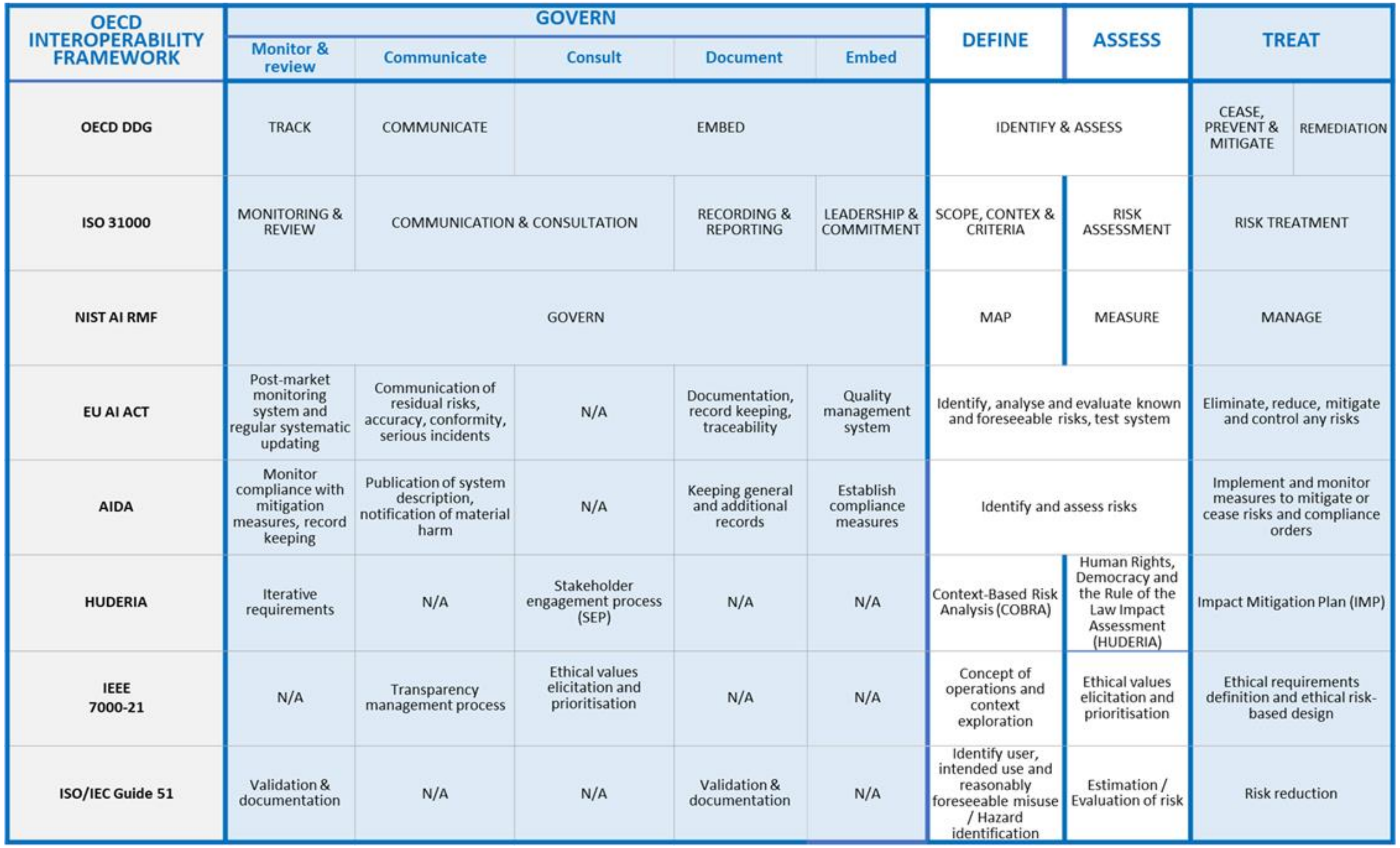
a) Structural view



b) Functional view

Source: (OECD, 2023[1])

Table 1. High-level mapping of selected risk management frameworks for AI to the Interoperability Framework



OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT	
	Monitor & review	Communicate	Consult	Document	Embed				
OECD DDG	TRACK	COMMUNICATE	EMBED			IDENTIFY & ASSESS		CEASE, PREVENT & MITIGATE	REMEDIATION
ISO 31000	MONITORING & REVIEW	COMMUNICATION & CONSULTATION		RECORDING & REPORTING	LEADERSHIP & COMMITMENT	SCOPE, CONTEX & CRITERIA	RISK ASSESSMENT	RISK TREATMENT	
NIST AI RMF	GOVERN					MAP	MEASURE	MANAGE	
EU AI ACT	Post-market monitoring system and regular systematic updating	Communication of residual risks, accuracy, conformity, serious incidents	N/A	Documentation, record keeping, traceability	Quality management system	Identify, analyse and evaluate known and foreseeable risks, test system		Eliminate, reduce, mitigate and control any risks	
AIDA	Monitor compliance with mitigation measures, record keeping	Publication of system description, notification of material harm	N/A	Keeping general and additional records	Establish compliance measures	Identify and assess risks		Implement and monitor measures to mitigate or cease risks and compliance orders	
HUDERIA	Iterative requirements	N/A	Stakeholder engagement process (SEP)	N/A	N/A	Context-Based Risk Analysis (COBRA)	Human Rights, Democracy and the Rule of the Law Impact Assessment (HUDERIA)	Impact Mitigation Plan (IMP)	
IEEE 7000-21	N/A	Transparency management process	Ethical values elicitation and prioritisation	N/A	N/A	Concept of operations and context exploration	Ethical values elicitation and prioritisation	Ethical requirements definition and ethical risk-based design	
ISO/IEC Guide 51	Validation & documentation	N/A	N/A	Validation & documentation	N/A	Identify user, intended use and reasonably foreseeable misuse / Hazard identification	Estimation / Evaluation of risk	Risk reduction	

Source: (International Organization for Standardization, 2018[5]; OECD, 2023[1]; OECD, 2018[6]; US NIST, 2023[7]; European Commission, 2021[8]; Council of Europe, 2023[9]; IEEE, 2021[10]; International Organization for Standardization, 2014[11]; Government of Canada, 2023[12])

1.1 OECD Due Diligence Guidance for Responsible Business Conduct (OECD DDG)

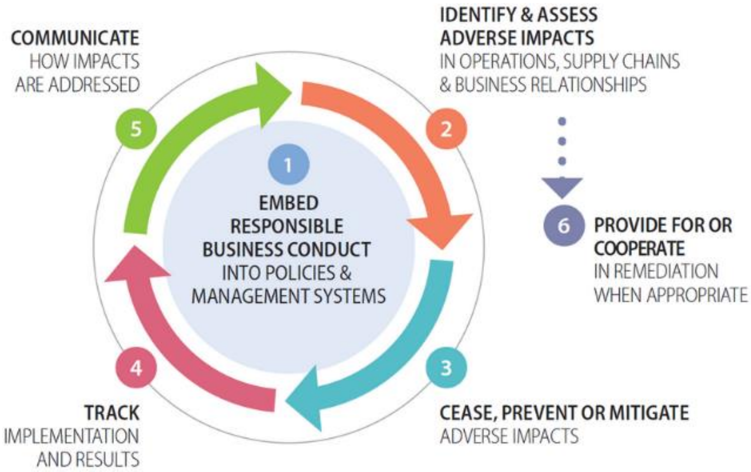
The OECD DDG is derived from the OECD Guidelines for Multinational Enterprises (MNE Guidelines). The MNE Guidelines are a set of government-backed voluntary recommendations for business to proactively address potential harms they may cause, contribute to, or are directly linked to. The MNE Guidelines specifically recommend that companies carry out due diligence to identify and address any adverse impacts associated with their operations, their supply chains or other business relationships. This approach to maximise the positive potential of business by first minimising the negative impacts forms the foundation for responsible business conduct (RBC).

Based on the recommendation in the MNE Guidelines that companies conduct due diligence to identify and address adverse impacts, the OECD has developed sector-specific guidance for carrying out supply chain due diligence in minerals, garment & footwear, agriculture and the financial sectors. Most recently, and most relevant to the discussion on new technology, the OECD has developed sector-agnostic OECD Due Diligence Guidance for Responsible Business Conduct (OECD DDG) that draws from and builds on sector-specific guidance but can be applied to all businesses in all sectors of the economy, including all companies in the AI value chain.

The framework in the OECD DDG consists of six top-level steps. The top-level steps are to (1) embed RBC into company policies and management systems, (2) identify and assess adverse impacts in operations, supply chains and business relationships, (3) cease, prevent or mitigate adverse impacts, (4) track implementation of efforts to address risk, (5) communicate on due diligence efforts and (6) provide for or cooperate in remediation when appropriate. These steps are meant to be simultaneous and iterative, as due diligence is an ongoing, proactive and reactive process (Figure 2. Graphical representation of the OECD Due Diligence Guidance for Responsible Business Conduct).

In particular, the OECD DDG specifies that when a company has caused or contributed to an impact, the company is expected to provide for or cooperate in remediation. Legitimate remediation mechanisms can include State-based or non-State-based processes through which grievances concerning business-related impacts can be raised and remedy can be sought.

Figure 2. Graphical representation of the OECD Due Diligence Guidance for Responsible Business Conduct


COMMUNICATE
HOW IMPACTS ARE ADDRESSED
5
IDENTIFY & ASSESS ADVERSE IMPACTS
IN OPERATIONS, SUPPLY CHAINS & BUSINESS RELATIONSHIPS
2
1
EMBED RESPONSIBLE BUSINESS CONDUCT
INTO POLICIES & MANAGEMENT SYSTEMS
6
PROVIDE FOR OR COOPERATE
IN REMEDIATION WHEN APPROPRIATE
4
3
TRACK
IMPLEMENTATION AND RESULTS
CEASE, PREVENT OR MITIGATE
ADVERSE IMPACTS

Source: (OECD, 2018[6])

Commonalities with the Interoperability Framework

Though the terminology and order of the OECD DDG high-level steps differ from the Interoperability Framework, the steps otherwise map very closely at both the top-level and in the more detailed recommendations (Table 2. Mapping top-level steps of the OECD DDG to the Interoperability Framework).

Differences with the Interoperability Framework

In terms of process, a notable difference between the OECD DDG and the Interoperability Framework is the role of remedy in the risk management process. In the Interoperability Framework, remedy is included throughout (but mainly under TREAT), while in the OECD DDG, remediation is included as a high-level category of its own. Additionally, the approach to GOVERN differs between the OECD DDG and the Interoperability Framework: in the OECD DDG, Consult and Document are mainly included under Embed.

A unique aspect of the OECD DDG and possible difference with the Interoperability Framework is the recommendation to address risks that the company might also contribute to or be directly linked to through business relationships. The Interoperability Framework and other standards discussed in this document focus on risks in the design, development and use of AI systems, but risk mitigation on the sale and distribution of products and services are less of a focus.

Table 2. Mapping top-level steps of the OECD DDG to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT	
	Monitor & review	Communicate	Consult	Document	Embed				
OECD DDG	TRACK	COMMUNICATE	EMBED			IDENTIFY & ASSESS		CEASE, PREVENT & MITIGATE	REMEDY

Source: (OECD, 2018[13]; OECD, 2023[1])

1.2 ISO 31000:2018 Risk Management – Guidelines (ISO 31000) and ISO/IEC 23894:2023

ISO 31000 provides general, sector-agnostic recommendations for managing any type of risk, including external risks to people and planet, but also internal risks to the value and reputation of an organisation. The overall objective of ISO 31000 is to be generic and applicable in many different circumstances, allowing for it to be customised to any organisation and its specific context. It structures its guidelines under general principles, a risk management process and a framework for leadership and commitment, all integrated into the same standard (Figure 3. Graphical representation of the ISO 31000:2018 Risk Management Guidelines).

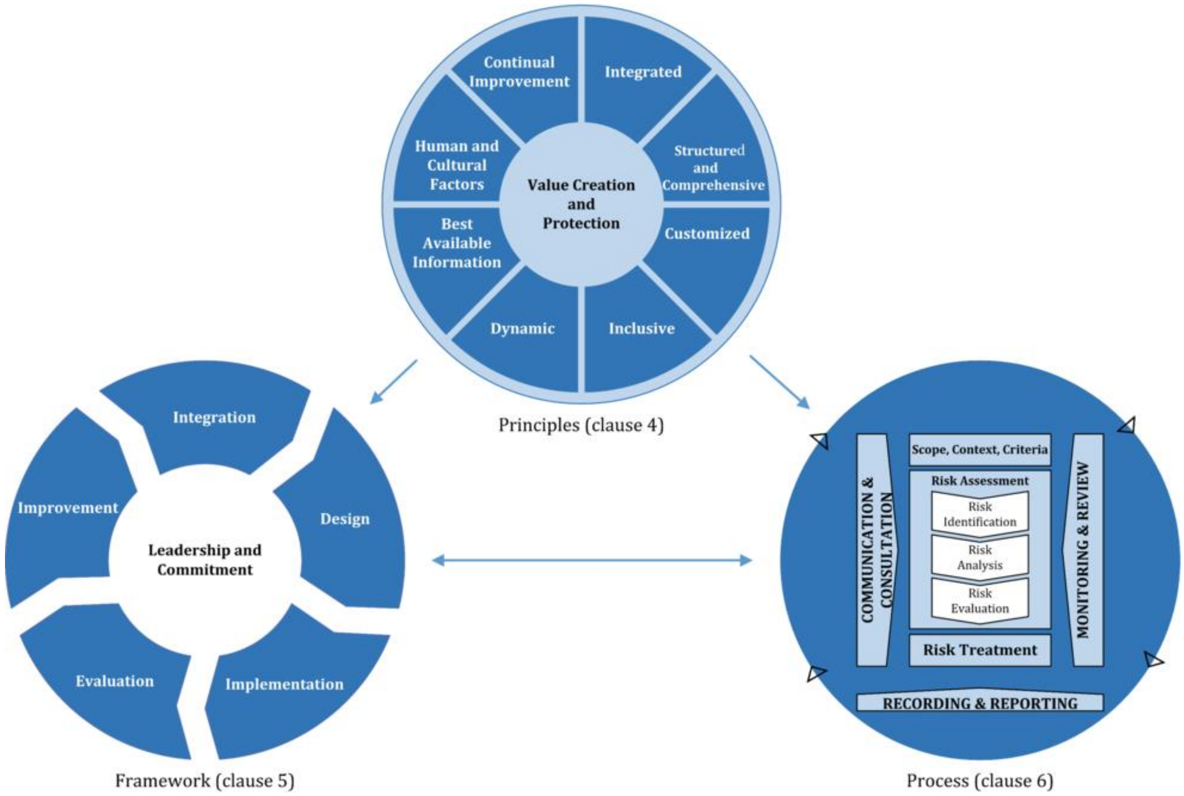
In February 2023, ISO and the International Electrotechnical Commission (IEC) developed a supplemental AI-specific Guidance to be used alongside ISO 31000, designated ISO/IEC 23894:2023. This document uses the same risk management framework as ISO 31000 and contextualises all AI-specific recommendations within the ISO 31000 structure. The AI-specific supplemental detail provided by ISO/IEC 23894:2023 specifically covers:

- Additional governance considerations regarding the development, purchase or use of an AI system.
- Stakeholder engagement for the purpose of improving human oversight, including with internal stakeholders, external impacted stakeholders, regulators.
- Tracking, record-keeping, and monitoring risk and risk management information.

- Re-assessments and re-evaluating risk management for continuous improvement over time.

At the top-level the two standards are the same, and for the purposes of this comparison exercise, they are considered together.

Figure 3. Graphical representation of the ISO 31000:2018 Risk Management Guidelines



Source: (International Organization for Standardization, 2018[5])

Commonalities with the Interoperability Framework

The ISO 31000 and the Interoperability Framework map robustly to one another at the top-level. The ISO 31000 risk management process recommends defining a risk scope, understanding context, conducting a risk assessment and treating the risks, all while recording, communicating and monitoring the risk management process. This is aligned with DEFINE, ASSESS and TREAT, and partially also with GOVERN. Some recommendations that map to GOVERN from the Interoperability Framework and other more general risk management principles such as embedding risk management at every level of company responsibility, stakeholder engagement and continuous improvement are also present in ISO 31000[6] (Table 3. Mapping top-level steps of ISO 31000 to the Interoperability Framework).

Differences with the Interoperability Framework

The main high-level difference between ISO 31000 and the Interoperability Framework relates to the 'Embed' function under GOVERN. In ISO 31000, Embed appears as a key element of the 'framework' (Figure 3, clause 5) and outside the core risk management 'process' (Figure 3, clause 6). Value creation and protection (Figure 3, clause 4), which includes some elements of DEFINE, is also separate from the core risk management process.

Additionally, ISO 31000 differs from the Interoperability Framework in that it considers risks and impacts to the organisation more narrowly. As a result, certain risk mitigation options (e.g., “do nothing further”, “retain risk” and “increase risk in order to pursue an opportunity”) may sometimes conflict with decreasing risks to people and planet. Similarly, because protecting the implementing organisation’s value is a core objective of ISO 31000, recommendations that would help ensure broader accountability (e.g., throughout the value chain) seem to be less of a priority.

The approach to risks and impacts in ISO/IEC 23894:2023 is broader and more consistent with the Interoperability Framework. It considers impacts to external stakeholders in more detail and recommends increased engagement with affected stakeholders and regulators to help ensue stronger human oversight of AI systems. Likewise, it recommends identifying how AI systems or components interact with pre-existing societal patterns that can lead to impacts on equitable outcomes, privacy, freedom of expression, fairness, safety, security, employment, the environment, and human rights broadly.

Table 3. Mapping top-level steps of ISO 31000 to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Consult	Communicate	Document	Embed			
ISO 31000	MONITORING & REVIEW	COMMUNICATION & CONSULTATION		RECORDING & REPORTING	LEADERSHIP & COMMITMENT	SCOPE, CONTEXT, & CRITERIA	RISK ASSESSMENT	RISK TREATMENT

Source: (International Organization for Standardization, 2018[5]; OECD, 2023[1])

1.3 NIST AI Risk Management Framework (NIST AI RMF)

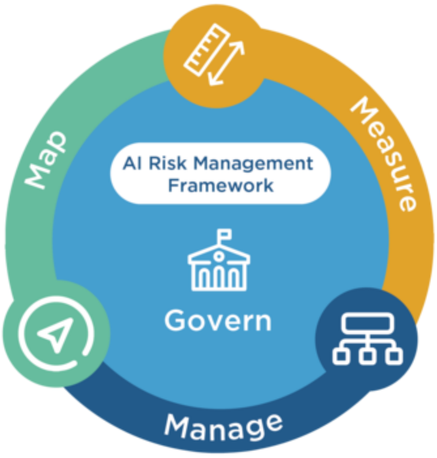
The NIST AI RMF is a voluntary framework, which aims to provide organisations with a process to help address risks throughout the AI lifecycle, with the objective of promoting trustworthy and responsible AI systems. It is intended to help manage both enterprise and societal risks related to the design, development, deployment, evaluation and use of AI systems. The NIST AI RMF is risk-based, outcome-focused and non-prescriptive. It defines the AI system lifecycle in line with OECD work (OECD, 2022[14]).

The NIST AI RMF refers to its top-level risk management process as “the Core”. The Core is composed of four functions (GOVERN, MAP, MEASURE and MANAGE) that are further broken down into categories and sub-categories. Like the other standards discussed, the four functions are non-sequential, continuous, iterative and meant to cross-reference one another (Figure 4. Graphical representation of NIST AI Risk Management Framework).

The GOVERN function focuses on policies, plans, organisation, responsibilities and accountability structures all focused on embedding AI risk management in all the organisation’s functions. The MAP function focuses on information gathering to establish visibility over the AI system lifecycle, AI capabilities, risks, benefits, impacts and stakeholders. The MEASURE function includes tracking metrics for trustworthy characteristics, social impact and human-AI configurations. The MANAGE function entails allocating resources to mapped and measured risks.

The NIST AI RMF covers three buckets of impacts: harm to people, harm to organisations and harm to ecosystems. This includes human rights, environmental and governance impacts.

Figure 4. Graphical representation of NIST AI Risk Management Framework



Source: (US NIST, 2023[7])

Commonalities with the Interoperability Framework

At a high-level, the NIST AI RMF is substantially aligned to the Interoperability Framework. DEFINE, ASSESS and TREAT are well aligned to the NIST AI RMF MAP, MEASURE and MANAGE steps and both frameworks include a GOVERN function (Table 4. Mapping top-level steps of the NIST AI RMF to the Interoperability Framework).

Differences with the Interoperability Framework

There are a few differences between the approach taken by the NIST AI RMF and the Interoperability Framework, especially regarding the GOVERN step. While the Interoperability Framework explicitly mentions monitoring, documenting, communicating and consulting as core cross-cutting components of an organisation's GOVERN function, NIST AI RMF's integrates these elements throughout the different steps of the risk management process. In contrast, the Embed function seems to be well contained within both frameworks' GOVERN function. While form may differ between both frameworks, content remains substantially similar.

In addition, some elements of DEFINE and ASSESS are included in the introductory sections of the NIST AI RMF (e.g., "Audience", "Risk framing" and "AI Risks and Trustworthiness") as opposed to under one or more of its core functions, namely GOVERN, MAP, MEASURE and MANAGE.

Table 4. Mapping top-level steps of the NIST AI RMF to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Consult	Communicate	Document	Embed			
NIST AI RMF	**GOVERN***					**MAP**	**MEASURE**	**MANAGE**

* Monitor; Consult; Communicate; Document are part of every high-level function of the NIST AI RMF and as such also feature in MAP, MEASURE and MANAGE.

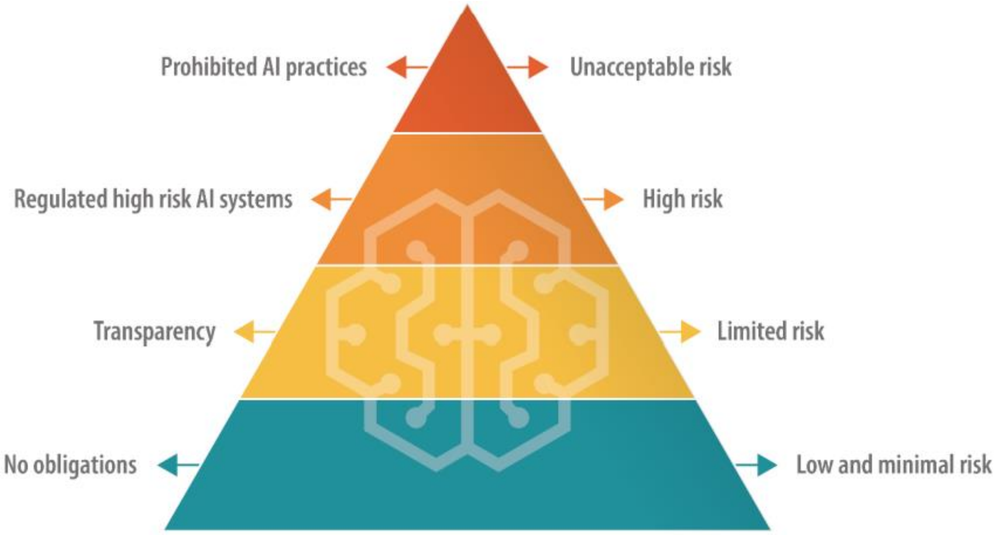
Source: (US NIST, 2023[7]; OECD, 2023[1])

1.4 European Union proposal for a regulation laying down harmonised rules on AI (EU AIA)

The EU AIA is a proposed binding legal framework requiring relevant companies to establish a risk-based approach to mitigate or prevent harms associated with certain uses of AI. It aims to be a flexible, horizontal framework, and to set minimum requirements to address AI-related risks without unduly constraining AI innovation. For the purposes of this report, the proposal offered by the European Commission will be analysed. Developers of high-risk AI systems will need to conduct both pre-deployment conformity assessments and post-market risk management to demonstrate that their systems meet all the requirements in the EU AIA's risk framework. It will apply to providers of AI systems, as well as to certain distributors, importers and users, subject to conditions. It will have a wide territorial reach, including to non-EU organisations that supply AI systems into the EU.

The EU AIA categorises different uses of AI as either entailing (i) unacceptable risk, (ii) high risk or (iii) low or minimal risk. Category (i) AI systems are prohibited for use or development (e.g. for subliminal distortion of a person's behaviour that may cause physical or mental harm; exploiting vulnerabilities of specific groups of people like the young, the elderly, or persons with disabilities; social scoring that may lead to unjustified or disproportionate detrimental treatment; and real-time remote biometric identification in publicly accessible spaces by law enforcement except for specific actions like searching for missing persons or counterterrorism operations). Category (ii) high risk AI systems are required to implement "risk management measures" among other conformity requirements (e.g., pertaining to data governance, disclosure, human oversight, record keeping, etc.). Category (iii) systems are subject to very few binding requirements but are encouraged to conform to voluntary codes of conduct (Figure 5. Graphical representation of the proposed EU AI Act risk classification). Under Article 65 of the EU AIA, market surveillance authorities may also evaluate certain AI systems that present a risk "to the health or safety or to the protection of fundamental rights" and require corrective action plans.

Figure 5. Graphical representation of the proposed EU AI Act risk classification



Source: (Madiega, 2022[15]) (European Commission, 2021[8])

Commonalities with the Interoperability Framework

Certain required risk management measures in the EU AIA map to the Interoperability Framework steps DEFINE, ASSESS and TREAT (Table 5. Mapping top-level EU AIA requirements to the Interoperability Framework). Namely, DEFINE and ASSESS: EU AIA Article 9(2)(a)-(c) requires providers of high-risk AI systems to identify and analyse known and foreseeable risks, estimate and evaluate what risks may arise from both the intended and reasonably foreseeable misuse of such systems and evaluate if they may pose any other risks. EU AIA Article 9(5) requires testing of high-risk AI systems to identify the most appropriate risk management measures.

TREAT: EU AIA Article 9(2)(d)-(7) sets out requirements for providers of high-risk AI systems to adopt suitable risk management measures to eliminate, reduce, mitigate or control any risks identified.

With regards to GOVERN:

- *Monitor & review*: EU AIA Article 9(2)(c) requires "evaluation of other possibly arising risks based on the analysis of data gathered from *the post-market monitoring system* referred to in Article 61". Article 9(2) states that "The risk management system shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring *regular systematic updating*".
- *Communicate*: EU AIA Article 9(4) requires residual risks of high-risk AI systems that are judged acceptable shall be "communicated to the user". Paragraph (49) requires that "the level of accuracy and accuracy metrics [of high-risk AI systems] should be communicated to the users". Article 9(4)(C) requires the "provision of adequate information" regarding risk estimation and evaluation. Section 5 AI requires providers of high-risk AI systems "to provide meaningful information about their systems and the conformity assessment carried out on those systems" and "inform national competent authorities about serious incidents or malfunctioning that constitute a breach of fundamental rights obligations". For non-high-risk AI systems, providers are required to provide information that flags the use of an AI system when interacting with humans.
- *Document*: The EU AIA contains documentation (including technical documentation), record-keeping and traceability requirements for high-risk AI systems (e.g., paragraph 46, 47, 54). Article 9(1) requires that the risk management system be documented[7]. Annex IV lists provisions for technical documentation.'
- *Embed*: The EU AIA contains requirements on human oversight that link closely to *Embed*, in particular Article 14 requires human oversight aimed "at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse..."

Differences with the Interoperability Framework

Some risk management measures that map to GOVERN in the Interoperability Framework are less clear and at present, distributed throughout the legislative proposal. For example, the EU AIA proposal does not seem to include consultation requirements with internal and external stakeholders and does not explicitly mention embedding risk management into organisational governance.

Regarding the latter, however, Article 17 requires providers of high-risk AI systems to put in place a "quality management system" to ensure compliance, including "an accountability framework setting out the responsibilities of the management and other staff". It mentions including the risk management system into the quality management system and as such, comes closest to embedding the risk management system into broader organisational governance.

In addition to Embed, the required quality management system includes other elements of governance such as monitoring and review, communication and reporting and documentation. A more detailed mapping between the EU AIA and the Interoperability Framework should account for the complementarities and overlaps between the risk management measures and the quality management system.

A unique feature of the EU AIA relative to the Interoperability Framework and the other standards described in this report, is that the regulator takes on certain top-level risk definition and assessment tasks by first making the determination as to what is considered a high-risk system. This provides a de-facto risk prioritisation mechanism for companies.

Table 5. Mapping top-level EU AIA requirements to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Consult	Communicate	Document	Embed			
EU AIA	Post-market monitoring system and regular systematic updating	N/A	Communication of residual risks, accuracy, conformity, serious incidents	Documentation, record keeping, traceability	Quality management system	Identify, analyse and evaluate known and foreseeable risks, test system		Eliminate, reduce, mitigate and control any risks

Source: (European Commission, 2021[8]; OECD, 2023[1])

1.5 Proposed Canada Artificial Intelligence and Data Act (AIDA)

The AIDA is a draft legislation proposed by the Government of Canada to regulate the design, development and deployment of AI systems. The AIDA takes a similar approach to the EU AIA by taking a risk-based approach to regulation. Under the AIDA, the following activities would be regulated:

- System design – including determining the AI system objectives and data needs, methodologies, or models based on those objectives.
- System development – including processing datasets, training systems using the datasets, modifying parameters of the system, developing and modifying methodologies, or models used in the system, or testing the system.
- Making a system available for use – deployment of a fully functional system, whether by the person who developed it, through a commercial transaction, through an application programming interface (API), or by making the working system publicly available.
- Managing the operations of a system – supervision of the system while in use, including beginning or ceasing its operation, monitoring and controlling access to its output while it is in operation, altering parameters pertaining to its operation in context.

Similar to the EU AIA, the AIDA identifies high-risk AI systems based on a number of criteria. It then requires entities in the AI system lifecycle to identify and address risks associated with those systems using a defined risk management framework.

Commonalities with the Interoperability Framework

Most of the top-level measures in the AIDA risk management framework align closely with the Interoperability Framework, specifically with regards to DEFINE, ASSESS and TREAT, as well as with some elements of GOVERN. (Table 6. Mapping top-level AIDA requirements to the Interoperability Framework).Under the legislation, covered companies are required to identify and assess risks, mitigate or cease risks, and monitor and document how risks are managed.

Differences with the Interoperability Framework

Certain elements of the Interoperability Framework are less clear at the top-level, but may added later as the legislation continues to take shape, namely under GOVERN (consult and embed). For example, while risk management procedures are required to be documented, it is unclear in what form they are expected to be shared with stakeholders and what role stakeholders would play in consulting on the risk management process. Likewise, while compliance procedures are required to be embedded in the covered companies, the legislation currently makes no mention of broader practices to embed risk management at every level of the decision making and AI development process within the company.

Table 6. Mapping top-level AIDA requirements to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Consult	Communicate	Document	Embed			
AIDA	Monitor compliance with mitigation measures, record keeping	N/A	Publication of system description, notification of material harm	Keeping general and additional records	Establish compliance measures	Identify and assess risks		Implement and monitor measures to mitigate or cease risks and compliance orders

Source: (Government of Canada, 2023[12])

1.6 Draft Council of Europe Human Rights, Democracy and the Rule of Law Risk and Impact Assessment (HUDERIA)

The HUDERIA, developed by the Council of Europe, represents a four-part approach to identify and address risks from AI systems. The issues in scope addressed is concrete and tied to the Council of Europe, EU and international standards on human rights, fundamental freedoms, elements of democracy and the rule of law. Each part of the HUDERIA risk management process flows chronologically, with the outcomes of each part feeding into the next. The four parts of the HUDERIA process can be implemented at the design, development and deployment phase.

- The process begins with a Context-Based Risk Analysis (COBRA) wherein the project team consolidates risk information about the project, identifies relevant stakeholders and develops a risk management plan.
- This is followed by the Stakeholder Engagement Process (SEP), where the outcomes of the COBRA are re-evaluated based on stakeholder feedback.
- Step three is the Human Rights, Democracy and the Rule of Law Impact Assessment (HUDERIA) where stakeholders and project teams come together to produce detailed evaluations of the potential and actual impacts that an AI system design, development and application could have on human rights and fundamental freedoms, democracy and the rule of law.
- Step four is the Impact Mitigation Plan (IMP) and related measures, which include assessing the severity of the potential adverse impacts; defining the measures to address these impacts; clarifying the roles and responsibilities of the various actors involved; monitoring impact mitigation efforts; and presenting remedy mechanisms.
- Finally, step five presents "iterative requirements" for parties to revisit the HUDERIA with a view on the dynamic and changing character of AI systems and the shifting conditions of the environments in which the systems are deployed.

Commonalities with the Interoperability Framework

The overall approach and expected outcomes of HUDERIA are consistent with the Interoperability Framework. In particular, COBRA, HUDERIA and the IMP map, respectively, to DEFINE, ASSESS and TREAT in the Interoperability Framework. Likewise, general principles on continuous improvement, monitoring, evaluation and stakeholder engagement are also present (Table 7. Mapping top-level steps of the HUDERIA to the Interoperability Framework).

Differences with the Interoperability Framework

The HUDERIA approach differs from the Interoperability Framework and the other standards described in this report in that it seems directed towards working-level risk management teams as they undergo the risk management process, rather than to the organisation as a whole. As such, elements from the Interoperability Framework relating to GOVERN seem to be lacking, such as communicating publicly on whether an AI system conforms to regulatory, governance, and ethical standards after assessing and treating risks and involving leadership to embed the risk management process across the organisational structure.

While documentation requirements are included through a note from the CoE Secretariat[8], they do not appear extensively in the January 2023 version of HUDERIA.

Table 7. Mapping top-level steps of the HUDERIA to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Consult	Communicate	Document	Embed			
HUDERIA	**Iterative requirements**	**SEP**	N/A	N/A	N/A	**COBRA**	**HUDERIA**	**IMP**

Source: (Council of Europe, 2023[9]; OECD, 2023[1])

1.7 IEEE 7000-21 Standard Model Process for Addressing Ethical Concerns during System Design (IEEE 7000-21)

The Institute of Electrical and Electronics Engineers (IEEE) developed the IEEE 7000-21 standard to help better integrate value-based considerations and stakeholder views in product or service innovation, design and modification phases. The standard's detailed recommendations are tailored narrowly to that specific stage of the AI system lifecycle. Essentially, IEEE 7000-21 helps define stakeholders and values, anticipate risks, engage stakeholders and integrate the outcomes of consultations into the innovation and design process.

The IEEE 7000-21 sets out processes, specific tasks, internal division of labour, recommended inputs and expected outcomes. Its top-level processes are described as follows:

(a) **Concept exploration stage**:

1. *Concept of Operations and Context Exploration Process,* to identify values and conduct a feasibility analysis. Includes understanding the ethical environment for system deployment and defining operational expectations.
2. *Ethical Values Elicitation and Prioritisation Process,* to consider ethical questions and priorities and to engage stakeholders. Includes defining and ranking ethical values to be implemented in system design and obtaining approval from management and other stakeholders.

(b) **Development stage**:

1. *Ethical Requirements Definition Process,* to develop risk mitigation strategies based on ethical value requirements.
2. *Ethical Risk-Based Design Process,* to translate ethical value requirements and risk management policies into implementable engineering targets.

(c) Transparency Management Process that extends through both stages to ensure appropriate communication with stakeholders in the design process.

Commonalities with the Interoperability Framework

While narrower and focused on certain aspects of the AI system lifecycle, the top-level processes and specific recommendations of IEEE 7000-21 are consistent with the Interoperability Framework. The outcomes of step (a.1) of the IEEE 7000-21 map to DEFINE by providing a description of the system's intended context of use, identifying stakeholders and collecting relevant legal, social, ethical and environmental contextual information. Step (a.2) maps to ASSESS specifically through risk prioritisation and the development of risk mitigation plans. Steps (b.1) and (b.2) map to TREAT through the development of risk mitigation strategies and implementable engineering targets (Table 8. Mapping top-level steps of the IEEE 7000-21 to the Interoperability Framework).

Step (c) maps to Communicate by covering communication with stakeholders throughout the design process.

Differences with the Interoperability Framework

As this standard is focused on innovation and design, it tackles risk mitigation slightly differently, by focusing on value setting and stakeholder engagement as a form of risk mitigation. Where the IEEE 7000-21 significantly differs from the Interoperability Framework is in the level of flexibility it offers in setting policies and determining risk scope. Whereas other frameworks more rigidly align with regulations and international norms in order to determine the risk scope and prioritisation, the IEEE 7000-21 offers detailed guidance on how ethical values can be leveraged to inform the overall risk management process.

Mapping some of the GOVERN functions from the Interoperability Framework to IEEE 7000-21 is challenging. Monitoring is not defined as a separate process but as an integral part of all IEEE 7000-21 steps. Documenting the risk management process and embedding it into organisational culture are noted at various points of IEEE 7000-21 but are not core focus areas.

Additionally, some high-level steps of the IEEE 7000-21 may not find a direct mapping with a step in the Interoperability Framework. For example, besides TREAT, translating ethical value requirements into engineering targets under step (b.2) *Ethical Risk-Based Design Process* also relates closely to ASSESS and Monitor & review in the Interoperability Framework. A second example is step (a.2) *Ethical Values Elicitation and Prioritisation Process*, which relates to ASSESS but also to Consult given its focus on stakeholder engagement.

Table 8. Mapping top-level steps of the IEEE 7000-21 to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Consult	Communicate	Document	Embed			
IEEE 7000-21	N/A	**Ethical values elicitation and prioritisation**	**Transparency management process**	N/A	N/A	**Concept of operations and context exploration**	**Ethical values elicitation and prioritisation**	**Ethical requirements definition and ethical risk-based design**

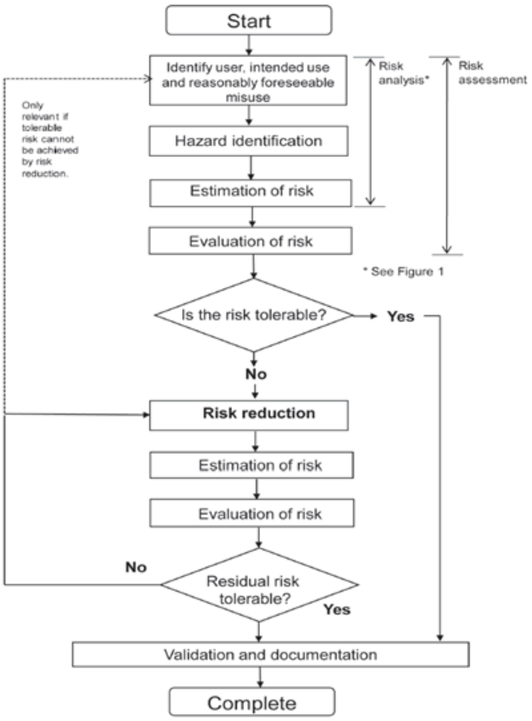
Source: (IEEE, 2021[10]; OECD, 2023[1])

1.8 ISO/IEC Guide 51:2014 3rd edition (ISO/IEC Guide 51)

ISO/IEC Guide 51 was developed by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) to inform the development of standards seeking to integrate product safety requirements in their risk management frameworks. It sets out how risks can be understood, in terms of severity and likelihood, and describes an iterative process to identify, estimate and reduce risks to "tolerable levels" (Figure 6. ISO/IEC Guide 51 Risk Assessment and Reduction Process). It also includes requirements for monitoring risk reduction efforts and documenting risk information. As this is a guide for people involved in the drafting of product safety standards, it does not provide specific detail or examples, but rather broader guiding principles on what type of product safety recommendations a standard should contain.

The harms considered broadly cover injury or damage to the health of people, or damage to property or the environment. ISO/IEC Guide 51 is aimed at safety risks arising from the design, production, distribution, use and destruction or disposal of products or systems, and considers minimising adverse impacts on the environment. The complete lifecycle of a product or system (including both the intended use and the reasonably foreseeable misuse) is considered. It also includes recommendations on whether a standard should include instructions for safe product use and safe testing, warning labels, and special packaging, where relevant.

Figure 6. ISO/IEC Guide 51 Risk Assessment and Reduction Process



Source: ISO/IEC Guide 51:2014

Commonalities with the Interoperability Framework

ISO/IEC Guide 51 focuses on risk identification, assessment and reduction. It is broadly consistent with DEFINE, ASSESS and TREAT in the Interoperability Framework, as well as some aspects of GOVERN such as monitoring and documenting.

Differences with the Interoperability Framework

The ISO/IEC Guide 51 differs from the Interoperability Framework in terms of scope and the type of harms covered. ISO/IEC Guide 51 does not include most of the sub-elements of GOVERN, such as embedding risk management policies, consulting with stakeholders and communicating risk management efforts (although ISO/IEC Guide 51 requirement to provide instructions for use and warnings could arguably be seen as a form of communication and documentation). Moreover, the cross-cutting nature of "Monitor & review" and "Document" in the Interoperability Framework (e.g., documenting process and outcomes at each step of the risk management process) is less evident under ISO/IEC Guide 51 "validation and documentation" step.

The scope of harms covered is also narrower than in the Interoperability Framework, focusing on injury or damage to health rather than on human rights violations more broadly. 'Human rights' are not mentioned in the ISO/IEC Guide 51.

Table 9. Mapping top-level steps of the ISO/IEC Guide 51 to the Interoperability Framework

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Consult	Communicate	Document	Embed			
ISO/IEC Guide 51	**Validation & documentation**	N/A	N/A	**Validation & documentation**	N/A	**Identify user, intended use and reasonably foreseeable misuse / Hazard identification**	**Estimation / Evaluation of risk**	**Risk reduction**

Source: (International Organization for Standardization, 2014[11])

2 Conclusions

In mapping risk management frameworks, the report finds general alignment between the Interoperability Framework at the top-level and different risk management frameworks. While the order of the risk management steps, the target audience, scope and specific terminology sometimes differ, all the frameworks follow roughly the same risk management process.

Yet while general approaches are aligned, high-level differences exist. Most of the differences relate to how the different frameworks approach the GOVERN function. For example, while in some frameworks governance activities are explicitly included under ‘GOVERN’, in others they are absent or distributed throughout the risk management process.

Some of these differences arise from the different scopes of different frameworks . For example, the OECD DDG considers risks associated with business relationships and recommends risk mitigation on the sale and distribution of goods, while ISO 31000 considers risks and impacts to the organisation more narrowly; NIST AI RMF covers harm to people, harm to an organisation and harm to an ecosystem; HUDERIA targets risks to human rights, fundamental freedoms, elements of democracy and the rule of law; the EU AIA and AIDA take a product safety approach to managing risks; and IEEE 7000-21 focuses on the integration of value-based considerations and stakeholder views into product or service design.

3 Next steps

Step 2. Analyse consistency, at both the conceptual and practical levels, of key concepts and terminology contained in different initiatives

The next step will be to take stock of commonalities and differences in the concepts and language/terminology of existing and emerging AI impact assessment and risk management frameworks **at the secondary levels** for each of the risk management stages. This step may also investigate alignment with relevant technical literature in terms of terminology and relationships between concepts. The objective is to provide a gap analysis on AI terminology and concepts that identifies and helps explain:

- Definitions and concepts that seem to generate significant consensus.
- Possibly incompatible or unclear components, which could hinder practical implementation of solutions, for example debates on the meaning and relationship of transparency, explainability and interpretability.
- A common understanding of the AI value chain; which actors are involved; and what risks exist at various stages in the value chain or with various AI use-cases.

Step 3. Translate analysis into good practice on due diligence for responsible business conduct in AI

A promising avenue to operationalise some of the risk management work would be to leverage the existing implementation and enforcement framework of the OECD DDG to develop good practices for responsible business conduct in AI that reflects and embeds AI and its specificities. This could be a timely and high impact contribution to advancing trustworthy, accountable AI.

This step proposes to bring together the Responsible Business Conduct and the AI risk management policy communities to align on terminology and frameworks. This could allow for a tailored approach to leveraging the high-level AI risk management interoperability framework (Figure 1. High-level AI risk management interoperability framework) in combination with the concepts from the OECD MNE Guidelines and DDG. Deliverables would include one or several workshops and a good practice report, handbook or FAQs, could clarify how Due Diligence Guidance principles for Responsible Business Conduct could be applied to AI.[9]

Step 4. Research and analyse the alignment of certification schemes with OECD RBC and AI standards

Increasing regulatory pressure and investor and consumer demand have led to a dramatic growth of certifications, standards and initiatives to address sustainability and environmental, social and governance issues in many sectors. In order to improve the quality, comparability and interoperability of certification standards and initiatives, the OECD has developed an alignment assessment process to evaluate the alignment of initiatives with the recommendations of OECD DDG.

This step could provide concrete recommendations to help translate and align AI practices into RBC good practices and vice versa leveraging, if relevant, an OECD alignment assessment framework and process that evaluate the alignment of initiatives with the recommendations of OECD DDG.[10]

Step 5. Develop an interactive online tool

This step would provide an interactive online tool to help organisations and stakeholders compare frameworks (see steps 1 and 2 above) and navigate existing methods, tools and good practices for identifying, assessing, treating and governing AI risks. Due diligence good practice, disaggregated by lifecycle actor and type of risk, would be linked to the Catalogue of Tools and Metrics for Trustworthy AI (https://oecd.ai/catalogue). The catalogue provides an interactive collection of the latest tools and resources available to help companies and other AI actors be accountable and ensure that AI systems and applications are trustworthy.

Annex A. Presentations relevant to AI risk from the OECD.AI network of experts

Since January 2022, the OECD.AI expert groups on classification and risk and on trustworthy AI have taken stock of key standards and initiatives in AI risk assessment and management (Table A.1. OECD.AI expert presentations

In September 2022, the two expert groups decided to join forces for both a work stream on AI risk and one on AI incidents. The co-chairs of the work stream on risk are Sebastian Hallensleben; Nozha Boujemaa and Andrea Renda.

Table A.1. OECD.AI expert presentations

a) OECD.AI Expert Group on Classification & Risk, January - September 2022

Name and date	Organisation	Presentation theme
Viknesh Sounderajah, 2 February 2022 (19th meeting)	Imperial College London	Forming AI Evidence Standards for Health Technology Assessment Programmes, presentation on the study of using the OECD framework for the classification of AI systems in the healthcare sector.
Mark Latonero and Elham Tabassi, 2 February 2022 (19th meeting)	National Institute of Standards and Technology (NIST)	Update on the development of the NIST AI Risk Management Framework.
Sebastian Hallensleben, 2 February 2022 (19th meeting)	CEN-CENELEC	Current European regulation/standardization aspects on AI risk assessment.
Kai Zenner, 24 March 2022 (20th meeting)	European Parliament	Overview of the JURI report's key proposed amendments to the EU AI Act.
Peter Deussen, 24 March 2022 (20th meeting)	ISO	Overview of ISO/IEC 23894's relevance for Artificial Intelligence risk management.
Aurelie Jacquet, 4 May 2022	Standards Australia Committee	Presentation on "Implementing AI Responsibly & Managing Risks, An Australian perspective"
Alpesh Shah, 4 May 2022	IEEE	Presentation on "IEEE 7000 and AI impact assessment"
Mark Latonero and Elham Tabassi, 22 September 2022	National Institute of Standards and Technology (NIST)	The NIST AI Risk Management Framework version 2.

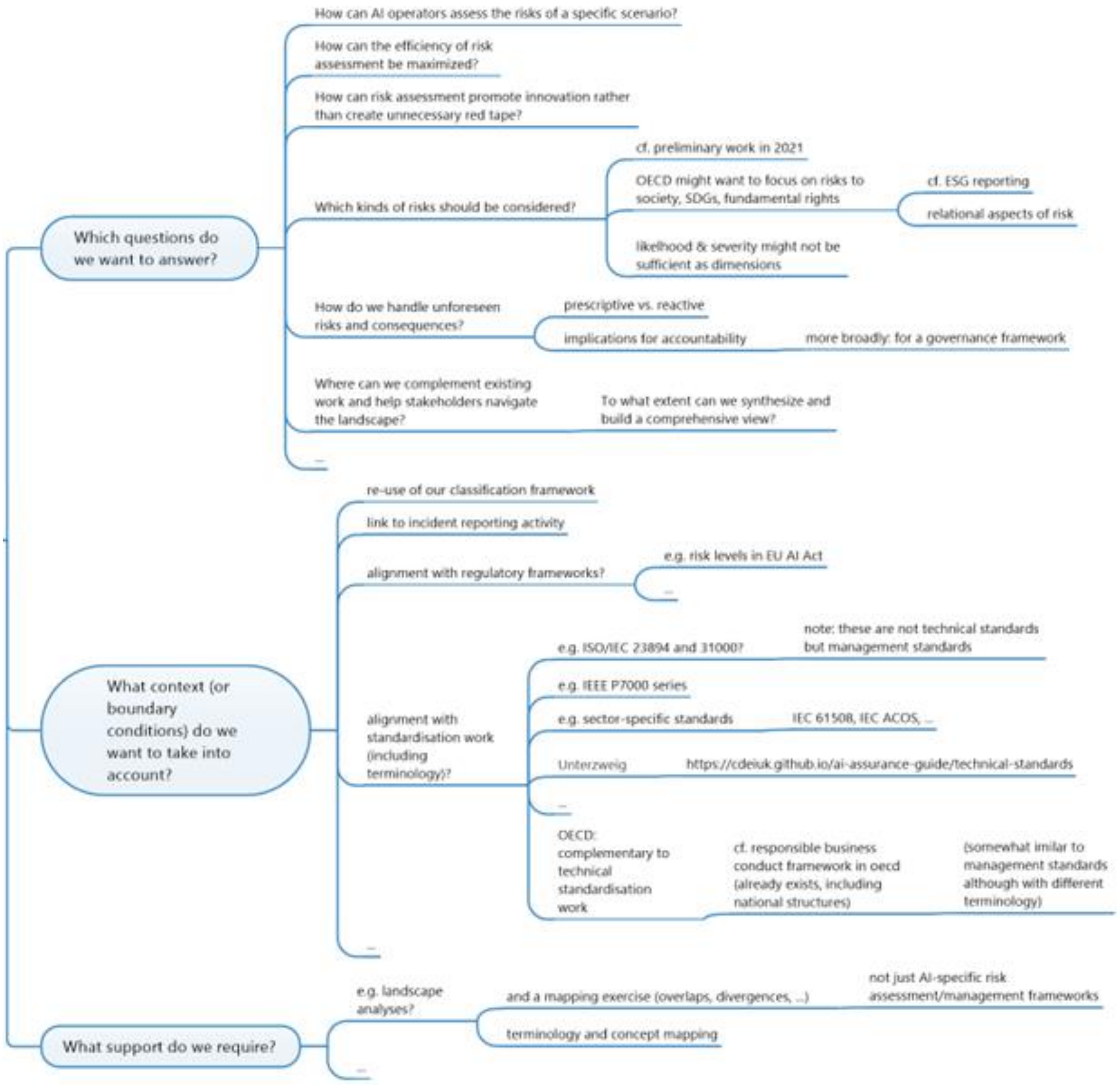
b) OECD.AI Expert Group on Tools & Accountability, June 2021 - September 2022

Name and date	Organisation	Presentation theme
Nozha Boujemaa, 25 June 2021 (11th meeting)	IKEA Retail (Ingka Group)	Algorithmic accountability, technical tools for accountability and value by-design models
Adriano Koshiyama and Emre Kazim, 16 July 2021 (12th meeting)	University College London (UCL)	Auditing algorithms from a technical perspective, including managing legal, ethical, and technological risks of AI, machine learning and associated algorithms
	Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE); German Research Center for Artificial Intelligence (DFKI)	Introduction to the AI projects at the Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE), including the "Trusted AI Initiative" that uses AI to optimise & certify AI

Ashley Casovan, 31 August 2021 (13th meeting) and 29 April 2022 (17th meeting)	Responsible AI Institute (RAI)	The Responsible AI Institute's work to design and develop a certification programme for responsible AI
Andrea Renda, 31 August 2021 (13th meeting)	Centre for European Policy Studies (CEPS)	Overview of CEPS' Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe and the FCAI Brookings/CEPS Forum for Cooperation on Artificial Intelligence
Craig Shank, 18 October 2021 (14th meeting)	Independent expert	The credibility of soft law to ensure accountability for artificial intelligence
Tyler Gillard and Rashad Abelson, 18 October 2021 (14th meeting)	OECD Centre for Responsible Business Conduct (RBC)	Responsible business conduct and accountability in AI and the link between the OECD AI Principles and the OECD Due Diligence Guidance
Clara Neppel, 13 January 2022 (15th meeting)	Institute of Electrical and Electronics Engineers (IEEE)	The IEEE 7000 Global Standard for addressing ethical concerns during system design
Vanja Skoric, 13 January 2022 (15th meeting)	The European Center for Not-for-Profit Law (ECNL)	Socio-legal architectures for sustainable AI development and the significance of human rights impact assessments (HRIA) as an instrument for accountability and trust
Stephanie Ifayemi, 13 January 2022 (15th meeting)	Digital Standards Policy, UK Department for Digital (DCMS)	The role of digital technical standards in the UK's National AI Strategy and the framework for G7 collaboration on digital technical standards
Jenny Brennan, 29 April 2022 (17th meeting)	Ada Lovelace Institute	An Ada Lovelace Institute project on algorithmic impact assessment in healthcare
Yordanka Ivanova, 12 July 2022 (18th meeting)	DG CONNECT, European Commission	An update on the EU AI Act
Yeong Zee Kin, 12 July 2022 (18th meeting)	Infocomm Media Development Authority of Singapore	An overview of Singapore's AI Verify initiative
Mikael Jansen, 16 September 2022 (19th meeting)	D-Seal - Danish Industry Foundation	D-Seal – a labelling program for IT security and responsible use of data in the EU
Kolja Verhage, 16 September 2022 (19th meeting)	Deloitte Risk Advisory	Lessons learned from implementation of values-based AI principles in the private sector

Annex B. Discussions by OECD.AI Expert Group on risk assessment considerations

Figure B.1. Discussions by OECD.AI Expert Group on risk assessment considerations


Which questions do we want to answer?
How can AI operators assess the risks of a specific scenario?
How can the efficiency of risk assessment be maximized?
How can risk assessment promote innovation rather than create unnecessary red tape?
Which kinds of risks should be considered?
cf. preliminary work in 2021
OECD might want to focus on risks to society, SDGs, fundamental rights
cf. ESG reporting
relational aspects of risk
likelihood & severity might not be sufficient as dimensions
How do we handle unforeseen risks and consequences?
prescriptive vs. reactive
implications for accountability
more broadly: for a governance framework
Where can we complement existing work and help stakeholders navigate the landscape?
To what extent can we synthesize and build a comprehensive view?
...
What context (or boundary conditions) do we want to take into account?
re-use of our classification framework
link to incident reporting activity
alignment with regulatory frameworks?
e.g. risk levels in EU AI Act
...
alignment with standardisation work (including terminology)?
e.g. ISO/IEC 23894 and 31000?
note: these are not technical standards but management standards
e.g. IEEE P7000 series
e.g. sector-specific standards
IEC 61508, IEC ACOS, ...
Unterzweig
https://cdeiuk.github.io/ai-assurance-guide/technical-standards
...
OECD: complementary to technical standardisation work
cf. responsible business conduct framework in oecd (already exists, including national structures)
(somewhat imilar to management standards although with different terminology)
...
What support do we require?
e.g. landscape analyses?
and a mapping exercise (overlaps, divergences, ...)
not just AI-specific risk assessment/management frameworks
terminology and concept mapping
...

Source: Presentation during meeting of OECD.AI Expert Group on AI Risk & Accountability (2022).

References

Council of Europe (2023), *Secretariat's Speaking Notes on HUDERIA for the Third Plenary meeting*. [9]

European Commission (2021), *Proposal for a Regulation laying down harmonised rules on artificial intelligence*, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence. [8]

Government of Canada (2023), *The Artificial Intelligence and Data Act (AIDA) – Companion document*, https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document. [12]

IEEE (2021), *7000-2021 - IEEE Standard Model Process for Addressing Ethical Concerns during System Design*, https://doi.org/10.1109/IEEESTD.2021.9536679. [10]

International Organization for Standardization (2018), *ISO 31000 - Risk Management*, https://www.iso.org/obp/ui#iso:std:iso:31000:ed-2:v1:en. [5]

International Organization for Standardization (2014), *ISO/IEC Guide 51:2014 Safety aspects — Guidelines for their inclusion in standards*, https://www.iso.org/standard/53940.html. [11]

Madiega, T. (2022), *Artificial intelligence act*, https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI%282021%29698792_EN.pdf. [15]

OECD (2023), *Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI*, OECD Publishing, Paris, https://doi.org/10.1787/2448f04b-en. [1]

OECD (2023), *Recommendation of the Council on Digital Security Risk Management*,, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0479. [16]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [3]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [14]

OECD (2019), *Recommendation of the Council on Artificial Intelligence*, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449. [4]

OECD (2019), *Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)*, https://doi.org/10.1787/d62f618a-en. [2]

OECD (2018), *OECD Due Diligence Guidance for Responsible Business Conduct*, https://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-Business-Conduct.pdf. [13]

OECD (2018), *OECD Due Diligence Guidance for Responsible Business Conduct*, http://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-Business-Conduct.pdf. [6]

US NIST (2023), *AI Risk Management Framework 1.0*, National Institute of Standards and Technology, Gaithersburg, MD, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. [7]

Notes

[1] Especially ISO 31000 and ISO 23894 on AI risk management that is targeted for publication in May 2023.

[2] Leveraging, for instance, the Catalogue of Tools and Metrics for Trustworthy AI https://oecd.ai/catalogue/).

[3] According to ISO 31000, risk is the "effect of uncertainty on objectives" and "an effect is a deviation from the expected. It can be positive, negative or both and can address, create or result in opportunities and threats." This work is concerned with the negative effects of risk.

[4] As defined in the EU AI Act proposal.

[5] The field includes major AI standardisation initiatives, including by the International Organization for Standardization (ISO), Institute of Electrical and Electronics Engineers (IEEE), International Telecommunication Union (ITU), National Institute of Standards and Technology (NIST), European Telecommunications Standards Institute (ETSI), Internet Engineering Task Force (IETF) and European Committee for Electrotechnical Standardization (CEN-CENELEC), with specific strands focusing on AI design (e.g. trustworthiness by design); AI impact, conformity and risk assessments; and risk-management frameworks for AI. It also includes governmental and intergovernmental initiatives such as the EU's proposal for a horizontal AI Regulation, the UK's AI Standards Hub, the European AI Alliance, the Council of Europe's Committee on Artificial Intelligence (CAI) and the EU-US Trade and Technology Council; certification schemes such as that of the Responsible AI Institute (RAII), the IEEE CertifAIEd and Denmark's D-Seal; and risk-management work to provide assurances for trustworthy AI through verification, validation and auditing.

[6] Clause 4 and 5.

[7] Linked to the EU AIA is the AI Liability Directive (ALD) which would create a rebuttable "presumption of causality" against any AI system's developer, provider, or user, and would make it easier for potential claimants to access information about specific high risk AI systems as defined by the EU AIA (category ii). Under the ALD, high risk AI systems could be required to disclose technical documentation, testing data and risk assessments subject to safeguards to protect sensitive information, such as trade secrets. Failure to produce such evidence in response to a court order would permit a court to invoke a presumption of breach of duty.

[8] "The Secretariat's view is that there should be an effective documentation protocol and flow of relevant information regarding the respective outcomes of each step of HUDERIA from the actor(s) mentioned

above to the competent domestic authority (or designated third parties) who should effectively supervise the process" (Council of Europe, 2023[9]).

[9] For example, the OECD is developing a Handbook on Due Diligence for Environmental Risks in Mineral Supply Chains, paired with the Due Diligence Guidance for Responsible Mineral Supply Chains. Other existing guidance documents in addition to the cross sectoral Due Diligence Guidance for RBC include sector-specific guidance in Agriculture, Garment & Footwear, Stakeholder Engagement in Extractives, Addressing Child and Forced Labour Risks in Mineral Supply Chains. FAQs have proven to be effective and appropriate for addressing narrow subject matters. See for example the FAQ on How to Address Corruption and Bribery Risks in Mineral Supply Chains.

[10] Guidelines for MNEs - Organisation for Economic Co-operation and Development (oecd.org)