

# ● Ethical and trustworthy Artificial Intelligence

BSI's introduction to the European Artificial  
Intelligence Draft Act (AIA)



 [Click on page numbers to skip to the content](#)

# Contents

3

Introduction

5

Definition of AI

7

Prohibited Artificial  
Intelligence Practices

Classification

8

Requirements

10

General Purpose AI Systems

11

Obligations

13

Conformity Assessment

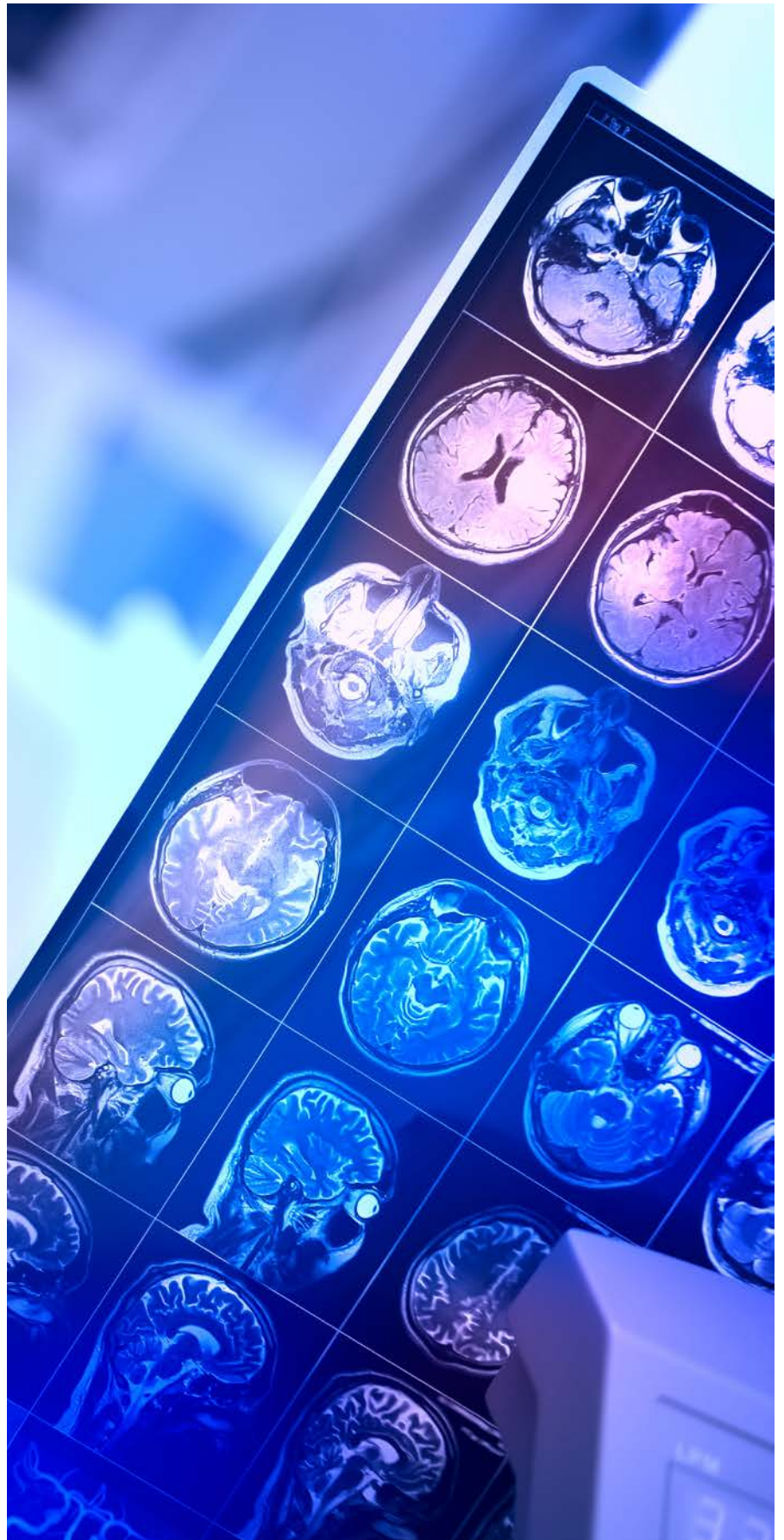
17

Role of Notified Bodies

Discussion

20

Bibliography



# Introduction

Artificial Intelligence (AI) is a fast-growing field. This is evidenced by the number of global publications, more than 334K in 2021 (Daniel Zhang, 2022), and the number of global AI patents, estimated above 142K in 2021 (Daniel Zhang, 2022).

As AI methods have already been implemented in different industries such as in the automotive sector through the use of AI Computer Vision and consumer products that harvest Natural Language Processing Algorithms (NLPs), considerations are raised on the Ethical and Trustworthy aspects of AI. BSI has witnessed the rapid growth in the implementation of AI in Medical Devices (MDs) and In Vitro Diagnostics (IVDs). Although mature and strict regulatory frameworks as well as standards exist for such industries, they were not designed to address AI specific challenges.

AI has demonstrated to perform similarly or even superiorly than non-expert humans, at specific tasks like image classification (Lu Yuan, 2021) (Daniel Zhang, 2022) and English language understanding (Alex Wang, 2019) (Daniel Zhang, 2022). AI is still far from being considered intelligent as it is currently task oriented (termed as narrow AI);

although there is no globally acceptable definition of the term “intelligence” (Shane Legg, 2007). However, this has not prevented humans from assigning human attributes to AI algorithms, such as being Ethical and Trustworthy.

Assurance services and conformity assessment bodies play an important role in placing trust in the output of a system. This can be illustrated in the trust patients place on a clinician's diagnosis. In emergency rooms, the patient is treated by a clinician that they have no prior knowledge of. It follows, therefore, that the patient has no knowledge of the experience or expertise of the treating physician. The clinician's cognitive process that leads to a decision is a “black box” to the patient. Why does the patient trust the clinical decision? Because the patient trusts the system - from the educational system that trained the clinician and validated their knowledge to the management of the hospital that makes sure results are reliably interpreted and that the equipment is maintained and calibrated for use in diagnosis. Additionally, without consciously knowing it, the patient trusts the system that sets the legal framework, develops standards



and so, implicitly, the Notified Bodies that grant access to MD/IVD devices to the EU market.

The European Union (EU) began its journey of an AI legislative framework in April 2019 by publishing “Ethics guidelines for trustworthy AI” (HLEG, 2019). This publication was followed by the Feb 2020 publication, a white paper on “AI-A European Approach to Excellence and Trust” (EC, White Paper on AI - A European approach to excellence and trust, 2020). In 2021 the European Commission released the draft AI Act (AIA) (EC, Artificial Intelligence Act, 2021). Following the initial release, the AIA text has changed and developed as it has made its way through the EU legislative process. In Nov 2022, the Commission adopted its position to take into negotiations with the European Parliament (Council, Nov 2022). The AIA is expected to become a regulation early in 2024, and is expected to come into application in 2027, according to the latest text, where the transition period is defined as 3 years. The legislative process for the GDPR (General

Data Protection Regulation) took 4 years, with an implementation period of 2 years (Floridi, 2021).

Once the AIA becomes law, as a regulation it will have immediate impact throughout the EU. AIA is expected to have a global effect, not because it is the first legal framework impacting AI - there are already other frameworks, standards and guidelines already published, such as in the US (HOR, 2020) and China (Graham Webster, 2017) (PRC, 2021) - but because it is not restricted to the geographical territory of the EU. The AIA clearly states (Article 2, 1a) that the scope covers “providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country” and “(1c) providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union”. A similar effect to the GDPR - which has seen widespread global adoption - is expected by the application of AIA.



## Definition of AI (Art.3)

The definition of AI is important as it will dictate the products that fall under the scope of AIA, as well as filtering out products using the term AI incorrectly. A global concern on the definition of AI has been raised, leading to a broad range of definitions in official documents. Some of which are presented in Table 1 - Definitions (TC260, 2021) that make use of the term “intelligence”. These are ambiguous however, as there is no globally accepted definition of intelligence (Shane Legg, 2007) (Wang, 2019).

AIA has approached, in the initial 2021 release (EC, Artificial Intelligence Act, 2021), the definition in a more straightforward manner by stating specific techniques for developing AI in Annex I. One could argue that some of the Annex I AI techniques and approaches are broadly defined (e.g., Statistical approaches, inductive (logic) programming), allowing space for methods not widely accepted as AI to fall under the AI definition. However, this approach is straightforward and specific in comparison to more general AI definitions found elsewhere (Table 1). As Wang (Wang, 2019) comments, the use of an AI definition that only considers what is established methodology would exclude new techniques and might be an obstacle to innovation.

The recent EU Council General approach (Council, Nov 2022), defines AI more broadly (see Table 1). However, in the proposal there are additional statements refining (paragraph 6) on what is considered AI. Paragraphs 6a and b further refine AI approaches:

- “In particular, for the purposes of this Regulation AI systems ..., using machine learning and/or logic- and knowledge based approaches ...”
- “A system that uses rules defined solely by natural persons to automatically execute operations should not be considered an AI system”

- “Machine learning approaches include for instance supervised, unsupervised and reinforcement learning, using a variety of methods including deep learning with neural networks, statistical techniques for learning and inference (including for instance logistic regression, Bayesian estimation) and search and optimisation methods.”
- “Logic- and knowledge based approaches include for instance knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning, expert systems and search and optimisation methods.”

It is important to note that, irrespective of the proposed definition, it is the responsibility of the conformity assessment body, when a third-party assessment process is required, to judge and challenge whether a proposed AI product falls under the definition. The ambiguity in definitions has been encountered in other sectorial legislations and additional guidelines were issued to provide more clarity; Medical Device Guidance documents, endorsed by the Medical Device Coordination Group (MDCG), are an example; MDCG 2022-5 (MDCG, 2022-5) provides guidance on borderline products between Regulation (EU) 2017/745 and Directive 2001/83/EC.



**Table 1: Definitions**

Source	Definition
WHO (2021)	An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
OECD (Recommendation of the Council on Artificial Intelligence, 2022)	An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
OECD (OECD framework for the classification of AI systems - public consultation on preliminary findings, 2021)	An AI model is a computational representation of real world processes, objects, ideas, people and/or interactions that include assumptions about reality. Core characteristics of this dimension include the model characteristics; how the system is built (e.g., using expert knowledge, machine learning or both); and how it is used (e.g. for which objectives and using what performance measures).
US House of Representatives (2020)	The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to: <ul style="list-style-type: none"> <li>a Perceive real and virtual environments</li> <li>b Abstract such perceptions into models through analysis in an automated manner</li> <li>c Use model inference to formulate options for information or action</li> </ul>
Gov.uk (National Security and Investment Act, 2021)	"Artificial Intelligence" means technology enabling the programming or training of a device or software to: <ul style="list-style-type: none"> <li>a Perceive environments through the use of data;</li> <li>b Interpret data using automated processing designed to approximate cognitive abilities;</li> <li>c Make recommendations, predictions or decisions; with a view to achieving a specific objective.</li> </ul>
China (TC260, 2021)	Artificial Intelligence: The simulation, extension or expansion of human intelligence by means of perceiving the environment, acquiring knowledge, derivation and deduction using computers or the equipment controlled by them.
AIA (EC, Artificial Intelligence Act, 2021)	Article 3: "Artificial Intelligence System" (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with. Annex I, Artificial Intelligence techniques and approaches referred to in Article 3, point 1: <ul style="list-style-type: none"> <li>a Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning</li> <li>b Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems</li> <li>c Statistical approaches, Bayesian estimation, search and optimization methods</li> </ul>
AIA (EC, Artificial Intelligence Act, 2021; Council, Nov 2022)	"Artificial Intelligence System" (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts.

# Prohibited Artificial Intelligence Practices

The EU legal framework was built around the notion that AI should be Ethical and Trustworthy. It is not therefore a surprise that AI practices that contradict basic human rights are considered prohibited. AIA (Council, Nov 2022)(Article 5) prohibits AI practices that might lead to physical or psychological harm (Fig. 1).

**Figure 1: Prohibited AI**

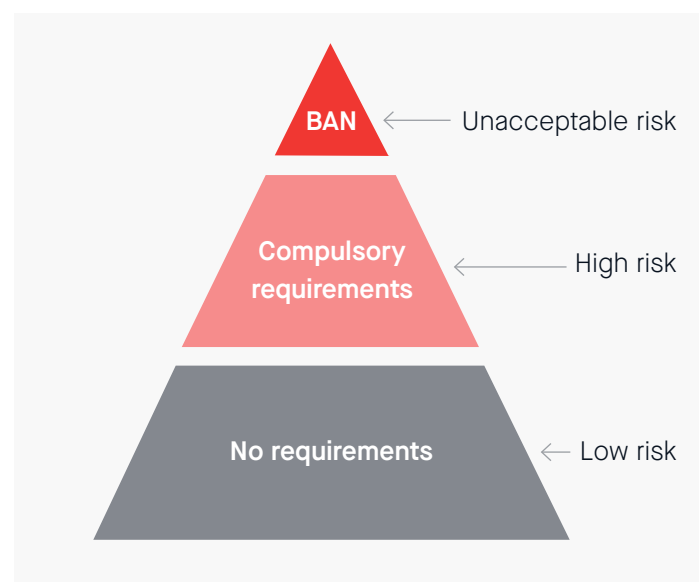


## Classification (Art. 6)

AIA (Council, Nov 2022) classifies AI system in two broad categories, High and Low Risk, described in annexes. When AI is a safety component or a product on its own and is covered under legislation set out in Annex II (Table 4) or described in Annex III (Table 4), then it is considered High-Risk AI. The exception is that, for AI described in Annex III, it will fall under high-risk category, “unless the output of the system is purely accessory”.

As a future failsafe mechanism, Article 7 empowers the Commission to update the Annex III list when two conditions are met: (a) the AI is intended to be used in areas covered by Annex III points 1-8; and (b) the AI poses a risk of harm to health, safety, or a risk to fundamental rights.

**Figure 2: Risk-Requirement pyramid**

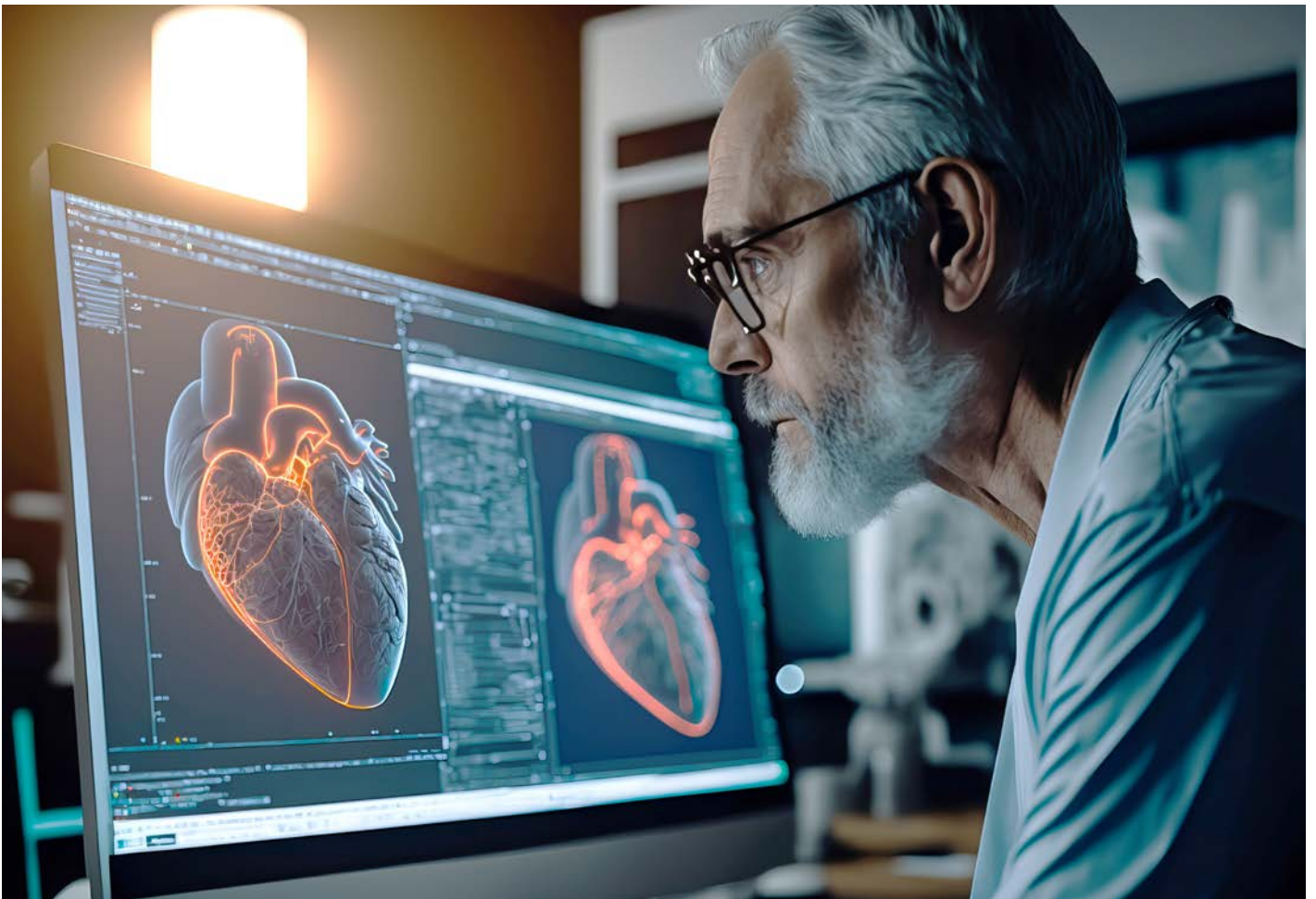


## Requirements (Ch. 2)

Requirements, conditions that a high-risk AI system should comply with, are set in Articles 9 to 15 (Council, Nov 2022). The description of requirements is at a high level and presumption of conformity with requirements is assumed when AI systems conform to Harmonized Standards (Art.40).

In the Council proposal (Council, Nov 2022) it is clarified that, when AI falls under a sectorial Union law (e.g., Medical Device Regulation - MDR), the

requirements of the AIA should be assessed under the conformity assessment process of the sectorial Union law. It is further clarified when high-risk AI systems are subject to obligations/requirements under relevant sectorial Union law, the AIA aspects may be part of the procedures/systems established pursuant to sectorial law. This is clarified for both Risk Management requirements (Art. 9 Sec 9) and Quality Management System obligations (Art. 17, Sec 2a).





**Table 2: AIA Requirements for High-Risk AI**

Requirement	Highlights
<b>Article 9: Risk Management (RM) System</b>	<ul style="list-style-type: none"> <li>• Providers should consider in RM known and foreseeable risks. Specific consideration should be given to impact on persons under the age of 18</li> <li>• Testing of AI systems in accordance with the intended purpose for the fulfilment of AIA requirements</li> <li>• Residual risks and overall residual risk should be judged as acceptable</li> </ul>
<b>Article 10: Data and data governance; Training, validation and testing data sets.</b>	<ul style="list-style-type: none"> <li>• Datasets and dataset methods should be subject to appropriate governance practices</li> <li>• Availability, quantity, and suitability of datasets should be assessed in advance</li> <li>• Training, validation and testing datasets should be free of errors (to the “best extent possible”) and take into consideration the specific context characteristics of the intended use context</li> </ul>
<b>Article 11: Technical Documentation</b>	<ul style="list-style-type: none"> <li>• When other legislation is applicable that requires Technical Documentation (Annex II, Sec A), a single documentation should cover both legal acts</li> </ul>
<b>Article 12: Record-keeping</b>	<ul style="list-style-type: none"> <li>• High risk AI should have automatic recording of events (logs) ensuring traceability throughout their lifetime to support post-market monitoring (PMS)</li> </ul>
<b>Article 13: Transparency and provision of information to users</b>	<ul style="list-style-type: none"> <li>• Operation should be sufficiently transparent to enable users to interpret the system’s output</li> <li>• Instructions for use (IFUs) in digital format. IFUs among other, should define intended use, performance characteristics (Article 15), foreseeable circumstances leading to risk, pre-determined changes, human oversight (Article 14), and expected lifetime</li> </ul>
<b>Article 14: Human oversight</b>	<ul style="list-style-type: none"> <li>• Appropriate human-machine interface for human oversight</li> <li>• Human oversight built into systems when feasible into systems and identified by the provider prior to placing on the market as appropriate</li> <li>• Users should be informed on automation bias, capabilities and limitations of AI, interpretation of outputs, and have the control of a decision not to use AI or terminate execution</li> <li>• For biometric identification products, before action is taken on AI output two natural persons should verify</li> </ul>
<b>Article 15: Accuracy, robustness, and cybersecurity</b>	<ul style="list-style-type: none"> <li>• Achieve an appropriate level of Accuracy, Robustness and Cybersecurity and performs consistently throughout their lifecycle</li> <li>• Accuracy and metrics disclosed in instructions for use (IFUs)</li> <li>• Resilience in errors within system or the environment in which AI operates</li> <li>• Robustness could be achieved by fail-safe backup plans</li> <li>• Evolving AI systems should ensure any possibly biased outputs are not used as inputs for future operations</li> <li>• Technical solutions to address data poisoning, adversarial examples or model flaws</li> </ul>

## General Purpose AI Systems (Art. 4a-c)

General purpose AI systems (GPAIS) are defined in AIA (Council, Nov 2022) in Art 3(1b) in an addition to the original text (EC, Artificial Intelligence Act, 2021) as:

“General purpose AI system” means an AI system that - irrespective of how it is placed on the market or put into service, including as open source software - is intended by the provider to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering, translation and others; a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems”

In Article 4a(2) it is clarified that requirements set out in Art. 4b shall apply irrespective of whether the general purpose AI is placed on the market or put into service as a pre-trained model.

GPAIS used as High-Risk AI systems or as components of such systems are required to comply in general with AIA requirements from the date of AIA application or 18 months after the entry into force. There will be implementing acts to specify AIA requirements to those GPAIS. The assessment process will follow internal control (Annex VI, points 3 and 4). When the instructions for use (IFUs) explicitly exclude all high risk uses, then the requirements of Art. 4b will not apply.

As it is a common practice for provider to use pre-trained models as the basis of fine tuning their final models, the adoption of these articles is expected to expand the applicability of AIA beyond the industries described in Annexes II and III (Table 4).



## Obligations (Ch. 3)

Chapter 3 of AIA (Council, Nov 2022) in Articles 16 to 29 set out a range of obligations of those involved in developing and supplying AI systems. Unusually for EU product legislation, the obligations are not only limited to Economic operators (Providers, manufacturers, Authorized Representatives, importers, distributors) but also to the users (Article 29). It has become clearer in the Council's text that users should have necessary competence, training, and authority for overseeing AI (Art. 29, Sec 1a) and those obligations should not apply to non-professional users (paragraph 58, Table 3 references some of the key obligations).



**Table 3: Overview of main obligation for High-Risk AI**

Obligations	Highlights
	Chapter 2, requirements
	Technical Documentation
	Logs automatically generated
	Conformity Assessment procedure
	Inform Competent Authorities (CA) and NBs for Non-compliance (NC)
	Affix CE marking
<b>Providers</b>	<p><b>QMS in place</b></p> <ul style="list-style-type: none"> <li>• Strategy for regulatory compliance</li> <li>• Procedures/systems for design/control/verification/development/validation/data governance and processing</li> <li>• Technical specifications</li> <li>• Risk Management System</li> <li>• Post Market Monitoring System</li> <li>• Incident reporting procedures</li> <li>• External communication handling</li> <li>• Resource management and accountability</li> </ul>

Obligations	Highlights
Economic Operators	<b>Authorized representatives appointed in the Union when no importer</b>
	<ul style="list-style-type: none"> <li>• Declaration of Conformity (DoC)</li> <li>• Documentation to comply with Chapter 2</li> </ul>
	<b>Importers</b>
	<ul style="list-style-type: none"> <li>• Ensure Conformity Assessment</li> <li>• Ensure Technical Documentation</li> <li>• Ensure CE</li> <li>• When AI presents risk: not place on the market, inform provider and authorities</li> <li>• Ensure storage/transport conditions</li> <li>• Communicate with CA</li> </ul>
<b>Distributors</b>	
<ul style="list-style-type: none"> <li>• Ensure CE mark and accompanying documentation</li> <li>• When AI presents risk: not place on the market, inform provider or importer</li> <li>• Ensure storage/transport conditions</li> <li>• Take Corrective Actions when needed</li> <li>• Capable of providing to authorities Documentation for conformity</li> </ul>	
<b>Manufacturers</b>	
Responsibility of AI compliance with AIA	
Users	<ul style="list-style-type: none"> <li>• Use in accordance with instructions for use (IFUs)</li> <li>• Human oversight indicated by provider</li> <li>• Input data control</li> <li>• Monitor AI operation</li> <li>• Maintaining logs</li> </ul>

# Conformity Assessment (Art.43)

Article 43 sets out the conformity assessment process of High-Risk AI systems. There are three proposed routes: internal control (Annex VI), third party assessment (Annex VII), and conformity assessment under other applicable legal acts for AI systems that fall under those (e.g., In Vitro Diagnostic AI devices falling under IVDR).

Substantial modification is described in the definitions (Art.3, Sec 23) (Council, Nov 2022), as a change that affects compliance with AIA requirements. Substantial modification of High-Risk AI systems leads to a requirement for a new conformity assessment procedure (paragraph 66). Evolving AI systems - those that continue to learn once placed on the market - require that they do so within pre-determined changes and such changes do not require re-assessment.

## AI systems listed in Annex III

High-risk Biometric Identification AI systems, listed in Annex III point 1 (Table 4), could follow either internal control (Annex VI) when Harmonized Standards or Common Specifications have been applied, or quality management system and Technical Documentation assessment by a Notified Body (Annex VII).

High-Risk AI systems listed in Annex III points 2 to 8 (Table 4) should follow internal control (Annex VI). No involvement of a Notified Body is required. However, the Commission is able to amend the legislation to require High-Risk AI systems of Annex III points 2-8 to undergo a conformity assessment of the quality management system and Technical Documentation by a Notified Body (Annex VII).

## AI systems listed in Annex II, Section A

For High-Risk AI falling under other EU legislation (Annex II, Sec A), that legislation - described in Annex II - is considered the "parent" legislation. For High-Risk systems listed in Annex II Section A (Table 4), providers should follow the conformity

assessment process dictated by the sectorial legislation. However, requirements set in AIA apply (Chapter 2, Table 2 above). Notified Bodies under the appropriate legal acts, should carry out the conformity assessment and ensure that this is performed by competent personnel (Article 33 point 10). If the applicable legal act does not require third party assessment (e.g., MDR Class I devices), and the provider has used Harmonized Standards or Common Specifications, they may use this option to not undergo third party assessment.

## AI systems listed in Annex II, Section B

There is no specific wording in Article 43 of the Council's text (Council, Nov 2022) on AI falling under Annex II Section B (Table 4). Article 2(2) describes that only articles 84 and 53 shall apply, insofar the AIA requirements have been incorporated under the applicable legislation.

Amendments to those legal acts are defined in AIA Articles 75 to 82, stating that requirements of AIA "shall be taken into account".

## Sandboxes (Art. 53 - 54b)

Article 53 covers AI regulatory sandboxes, which may be established by National Competent Authorities. Sandboxes may be used before systems are placed on the market or put into service, including testing in real word conditions for AI systems described in Annex III (Art. 54a), and are described as a controlled environment for fostering innovation (paragraph 72). However, it is not clear how sandboxes will interplay with conformity assessment processes.

- Art. 53, Sec (1b)(b): “**facilitate and accelerate access** to the Union market for AI systems, in particular when provided by small and medium enterprises (SMEs), including start-ups”
  - Art. 53 (4a):  
“... the national competent authority shall provide, where applicable, a written proof of the activities successfully carried out in the sandbox ... Such written proof and exit report could be taken into account by market surveillance authorities or notified bodies, as applicable, **in the context of conformity assessment procedures or market surveillance checks**”
  - Recitals paragraph 72: “The supervision of the AI systems in the AI regulatory sandbox should therefore cover their development, training, testing and validation **before the systems are placed on the market or put into service**, as well as the notion and occurrence of substantial modification that may require a new conformity assessment procedure.”
- SMEs including startups shall be prioritized for participation in sandboxes (Art. 55, Sec 1b). This is not the only support offered for SMEs in the AIA. Additional wording on fees (Art. 55, Sec 2), on QMS (Art. 55a, Sec 1), on requirements set by article 4b on IFUs excluding all high risk uses (Art. 55a, Sec 3), on fines (Art. 71, sec 3, 4 and 5), and on Technical Documentation (Art. 11, sec 1) can be found.

**Table 4: Annexes II and III; High-Risk AI systems**

Annex II, Section A	Annex II, Section B	Annex III
Directive 2006/42/EC Machinery Directive 2009/48/EC Toys Directive 2014/34/EU Protective Equipment in Explosive Atmospheres Directive 2013/53/EU Craft Directive 2014/33/EU Lifts Directive 2014/53/EU Radio Equipment Directive 2014/68/EU Pressure Equipment Regulation (EU) 2016/424 Cableway Installations Regulation (EU) 2016/425 PPE Regulation (EU) 2016/426 Appliances Gaseous Fuels Regulation (EU) 2017/745 Medical Devices Regulation (EU) 2017/746 In Vitro Diagnostics	Regulation (EC) No 300/2008: Aviation security Regulation (EU) No 168/2013 2-3 Wheel Vehicles Regulation (EU) No 167/2013 Agricultural vehicles Directive 2014/90/EU Marine equipment Directive (EU) 2016/797 Int. Rail System Regulation (EU) 2018/858 Market Surveillance OF Motor Vehicles Regulation (EU) 2019/2144 Type Approval Motor Vehicles Regulation (EU) 2018/1139 Civil Aviation	Access to and enjoyment of essential private and public services and benefits Law enforcement Education and vocational training Biometric Identification Employment, workers management and access to self-employment Critical infrastructure Migration, asylum and border control management Administration of justice and democratic processes

# Role of Standards and Common Specifications (Art. 40-41)

Chapter 5 of AIA (Council, Nov 2022) sets out the role of Harmonized Standards (HSs) and Common Specifications (CSs). Conformity with HSs and CSs is considered to presume conformity with AIA Chapter 2 requirements.

In 2021, the EU Science Hub published an AI landscape analysis (JRC, 2021). In this report the standards presented in table 5 were identified as the group of core standards (sec 8.4.5 of (JRC, 2021). Currently (Jun 2022) ISO/IEC JTC 1/SC 42 Artificial intelligence Technical Committee has published 16 ISO standards, while 25 are under development (Dec 2022).

In a recent draft standardization request (EC, Draft standardisation request to the European Standardisation Organizations in support of safe and trustworthy artificial intelligence, 2022) for



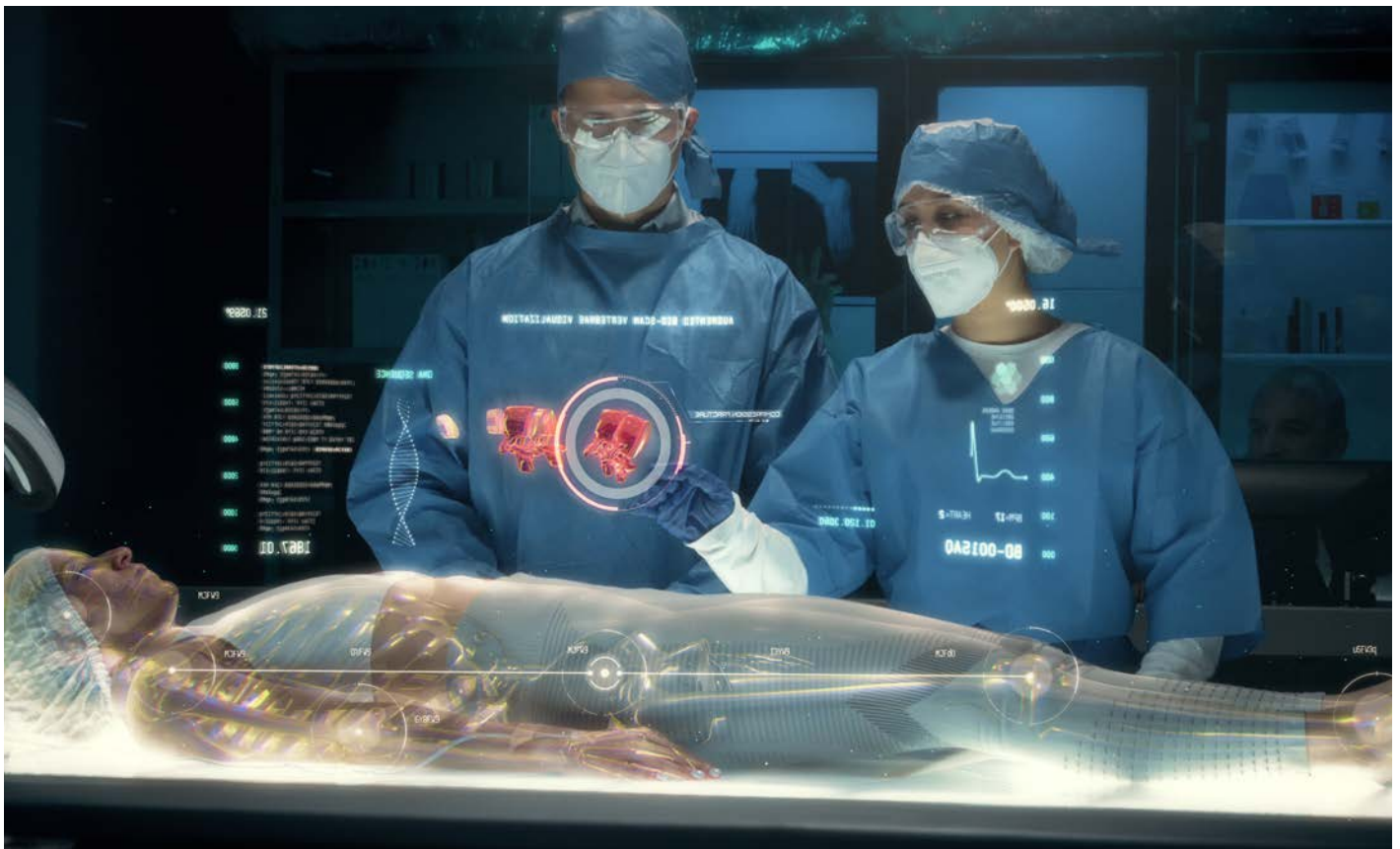
the AIA, the deadline for adoption by CEN and CENELEC of standards (found in Annex I of the request) is set to 31 Jan 2025. Table 1 of Annex I of this standardization request includes 8 AI standards on: risk management, governance and quality of datasets, record keeping, transparency, human oversight, accuracy, cybersecurity and robustness specifications, quality management systems and deliverable(s) on conformity assessment of AI systems.

The same document clarifies the interplay of standards between more than on applicable legislations:

- **Annex II, Section 2.1** on Risk Management System for AI systems:  
“Specifications shall be drafted in such a way that, for AI systems which are safety components of products, the risk management system aspects related to the AI system should, when applicable, be integrated into the risk management system for the overall product”
- **Annex II, Section 2.9** on Quality Management System for providers of AI systems, including post-market monitoring process:  
“Specifications shall be drafted such that the quality management system aspects related to the AI system may be integrated in the overall management system of the provider”

**Table 5: Core Standards (JRC, 2021)**

Standards	Title	Dec 2022 status
ISO/IEC 4213	ISO/IEC DTS 4213.2 Information technology - Artificial Intelligence - Assessment of machine learning classification performance	Published
ISO/IEC 5338	ISO/IEC CD 5338 Information technology - Artificial intelligence - AI system life cycle processes	40.20 DIS ballot initiated: 12 weeks
ISO/IEC 23894	ISO/IEC DIS 23894 Information technology - Artificial intelligence - Guidance on risk management	60.00 International Standard under publication
ISO/IEC 24027	ISO/IEC TR 24027:2021 Information technology - Artificial intelligence (AI) - Bias in AI systems and AI aided decision making	Published
ISO/IEC 38507	ISO/IEC 38507:2022 Information technology - Governance of IT - Governance implications of the use of artificial intelligence by organizations	Published
ISO/IEC 42001	ISO/IEC CD 42001.2 Information Technology - Artificial intelligence - Management system	Published



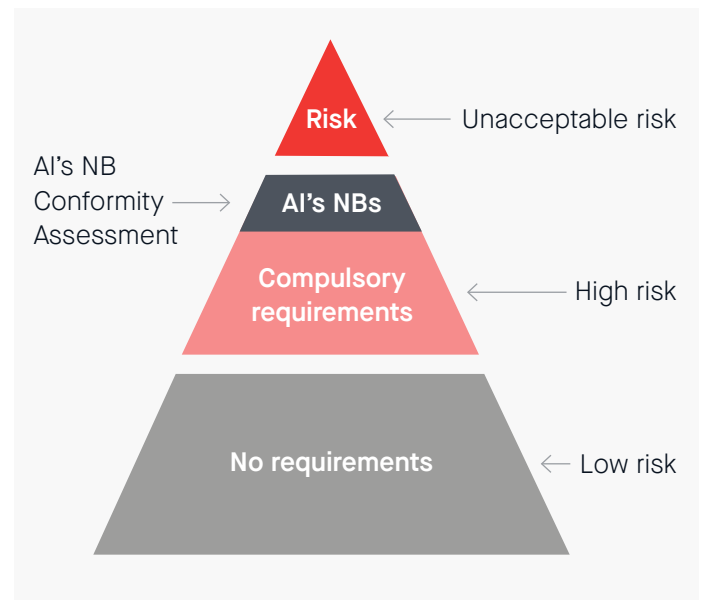


# Role of Notified Bodies

The third party QMS and Technical Documentation assessment process is described in Annex VII, valid for devices listed in Annex III, point 1 (Biometric identification, Table 4). Article 33 of AIA dictates requirements of Notified Bodies (NBs) under AIA. Some of the important requirements of AI NBs are the presence of a quality system, resources and competence, liability insurance and impartiality. The role of the AI NBs is currently limited to the Biometric Identification devices (Annex III, point 1), but this could be subject to change according to AIA Article 7.

AI High-Risk products or safety components that fall under the legal acts listed in Annex II Section A (Table 4) should undergo a conformity assessment process dictated by the applicable sectorial Union law by the appropriate NB. To illustrate this further, a Medical Device that falls under MDR (Regulation (EU) 2017/745) is subject to third party assessment by a NB designated under MDR, according to the conformity assessment route chosen by the manufacturer in line with the device classification. However, requirements for AI NBs are also valid for NBs under those legal acts. This is set out in Art. 43(3), where NBs under the sectorial Union legislation should be designated for AIA requirements laid down in Art. 33(4) Independence, (9) competence and (10) permanent availability of competent personnel.

**Figure 3: Artificial Intelligence Notified Bodies involvement in AI product third party assessment**



## Discussion

The application of the AIA is expected to have a global effect as it has a horizontal applicability across industries, it covers all providers and users where the output of the AI system is used in the EU, and standardization creates harmonization increasing business opportunities for large markets such as the EU.

Although the majority of industries do not fall directly under the requirements of the proposed AIA, a domino effect is expected with the introduction of requirements for General Purpose AI Systems.

This will further increase the regulatory burden for companies that do not operate under a strict regulatory framework and their current operation does not include oversight by auditing authorities.

Requirements set by the AIA are high-level, to be refined by Harmonized Standards and Common Specifications. Some of the requirements, such as human oversight, might be difficult to implement in high volume/speed applications. The AIA mitigates this by including wording that suggests human oversight will be an obligation assigned to professional rather than lay users of AI.

The interplay between AI standards and standards applicable to other sectorial Union law has been clarified for a small number of standards. As an example, how will the Medical Devices ISO 14971 Risk Management (RM) standard work with ISO/IEC 23894 AI RM? Both the AIA (Council, Nov 2022) as well as the draft request for standardization (EC, Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence, 2022) address this question by stating that aspects of AI will be intergraded into procedures/system applicable to sectorial Union law.

Incident reporting (Art. 62) requires providers of High-Risk AI systems to report to the market surveillance authorities. However, it is clarified that for other regulations reporting is also applicable, such as MDR 2017/745 and IVDR 2017/746. Notification according to AIA will be limited to serious incidents relevant to breach of obligations to protect fundamental rights (Art. 3, Sec 44c).



The definition between those standards of an incident is not the same, which further complicates the reporting process:

- MDR Art. 2(58): “Serious adverse event” means any adverse event that led to any of the following:
  - a Death
  - b Serious deterioration in the health of the subject, that resulted in any of the following:
    - Life-threatening illness or injury
    - Permanent impairment of a body structure or a body function
    - Hospitalisation or prolongation of patient hospitalisation
    - Medical or surgical intervention to prevent life-threatening illness or injury or permanent impairment to a body structure or a body function
    - Chronic disease
  - c Foetal distress, foetal death or a congenital physical or mental impairment or birth defect
- AIA Art. 3(44): “Serious incident” means any incident or malfunctioning of an AI system that directly or indirectly leads to any of the following:
  - a The death of a person or serious damage to a person’s health
  - b A serious and irreversible disruption of the management and operation of critical infrastructure
  - c Breach of obligations under Union law intended to protect fundamental rights
  - d Serious damage to property or the environment

Sandboxes is another concept introduced by the AIA. Although sandboxes are meant to foster innovation, and specific actions are to be taken for SMEs and start-ups, it is not clear what their role will be in the assessment process. Further clarifications are required on the role of sandboxes in the assessment process.

Predetermined changes apply to evolving AI. However, there is no definition on what predetermined changes are, or what should be included in the initial assessment process and monitored in the post market surveillance phase for this purpose.

Although there are other areas in the AIA that require further clarification, it is not the intent of this paper to shortlist all the areas, but rather to provide a brief introduction to the reader of the AIA. It is clear that the AIA is a necessary development as remaining idle in an era where AI is already in place is not an option. Over time, explanatory and supplementary documentation will need to be developed to provide additional clarity to support interpretation of the legal text.

The horizontal approach of the AIA is under global debate, as other legislators believe a vertical

approach, amending existing industry specific legislation, is more appropriate because it takes into consideration context specific AI risks and requirements. Once the EU adopts AIA, it will differentiate conformity assessment process to other countries, like the UK. Having a global process for assessing AI would be beneficial to AI providers, as a single application would cover multiple jurisdictions. There have been initial actions taken to this direction, however, such divergence in approaches already exists under most other product legislation.

#### **DISCLAIMER**

The current paper is the Author's understanding and interpretation of the draft Artificial Intelligence Act. As the AIA is still a draft, it is likely to be subject to changes prior to becoming an EU Regulation.



# Bibliography

Alex Wang, Y. P. (2019). SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems. Vancouver: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019).

Comission, E. (2022, Dec). Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence. Retrieved from here.

Council, G. S. (Nov 2022). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Council of the EU.

Daniel Zhang, N. M. (2022). The AI Index 2022 Annual Report. AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University.

EC. (2020). White Paper on AI - A European approach to excellence and trust. Brussels: European Commission.

EC. (2021, Apr). Artificial Intelligence Act. Proposal for a regulation of the European Parliament and the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Brussels: European Commission.

EC. (2022). C(2022) 546 ANNEX to the Commission Notice; The 2022 annual Union work programme for European standardisation. Brussels: European commission.

Floridi, L. (2021). The European Legislation on AI: a Brief Analysis of its Philosophical Approach. *Philosophy and Technology*, 34, pages 215-222.

Gov.uk. (2019, Jun 10). Guidance: A guide to using artificial intelligence in the public sector. Retrieved from gov.uk

Gov.uk. (2021). National Security and Investment Act. Retrieved from here

Graham Webster, R. C. (2017). A Next Generation Artificial Intelligence Development Plan. Retrieved from New America

HLEG, H.-L. E. (2019). Ethics Guidelines for Trustworthy AI. Brussels: European Commission.

HOR, H. O. (2020). WILLIAM M. (MAC) Thornberry National Defense Authorization Act For Fiscal Year 2021;. Washington: U.S. Government Publishing Office.

JRC. (2021). AI Watch: AI Standarisation Landscape. Luxembourg: Joint Research Centre (JRC), the European Commission's science and knowledge service.

Lu Yuan, D. C.-L. (2021). Florence: A New Foundation Model for Computer Vision. *Computer Vision and Pattern Recognition (arXiv:2111.11432)*, 1-17.

MDCG. (2022-5). Guidance on borderline between medical devices and medicinal products under Regulation (EU) 2017/745 on medical devices. Medical Device Coordination Group Document.

OECD. (2021). OECD Framework for the Classification of AI Systems - Public Consultation on Preliminary Findings. Organisation for Economic Co-operation and Development.

OECD. (2022). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.

PRC. (2021). A new generation of artificial intelligence ethics code.

Shane Legg, M. H. (2007). A Collection of Definitions of Intelligence.

TC260. (2021). Guidelines for the Practice of Cybersecurity Standards - Guidelines for the Prevention of AI Ethical Security Risks. The National Information Security Standardisation Technical Committee of China.

Wang, P. (2019). On Defining Artificial Intelligence. Journal of Artificial General Intelligence, 10(2):1-37.

WHO. (2021). Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. ISBN 978-92-4-002920-0: World Health Organization.

## **DISCLAIMER**

All rights reserved. Copyright subsists in all BSI publications, including, but not limited to, this white paper. Except as permitted under the Copyright, Designs and Patents Act 1988, no extract may be reproduced, stored in a retrieval system or transmitted in any form or by any means - electronic, photocopying, recording or otherwise - without prior written permission from BSI. While every care has been taken in developing and compiling this publication, BSI accepts no liability for any loss or damage caused, arising directly or indirectly in connection with reliance on its contents except to the extent that such liability may not be excluded in law.

### BSI UK Approved Body (0086)


Kitemark Court, Davy Avenue, Knowlhill  
Milton Keynes MK5 8PP  
United Kingdom

 +44 345 080 9000

 eu.medicaldevices@bsigroup.com

### BSI Netherlands Notified Body (2797)

Say Building, John M. Keynesplein 9  
1066 EP Amsterdam  
The Netherlands

 +31 20 346 0780

 eu.medicaldevices@bsigroup.com



Read more about our  
certification services  
on our website

[bsigroup.com/medical](https://www.bsigroup.com/medical)



Find us on LinkedIn