EC2N20

Position de Confiance.ai par rapport à l'AI Act

contact@confiance.ai | www.confiance.ai

Document reference: 220A

1
Contributors

	Name	Organisation	Role
Responsible for the deliverable	Henri Sohier	IRT SystemX	
Scientific responsible	Serkan Odabas	Valeo	
Co-authors	Bertrand Braunschweig	Confiance.ai	
	Morayo Adedjouma	CEA	
	Romane Vernhes	Sopra Steria	(affectée Confiance.ai)
	Nicolas Winckler	Eviden	
	Marina Bojarski	Eviden	(contributrice hors Confiance.ai)
	Joseph Machrouh	Thales	
	Hatem Hajri	Safran	
	Georges Jamous	Airbus Protect	
Reviewers	Kevin Mantissa	IRT SystemX	
	Karla Quintero	IRT SystemX	
	Eric Jenn	IRT Saint Exupery	
	Xavier Leroux	Thales	

2 Document control

Revision	Date	Commentary	Author
0.1	13/05/2024	Squelette	Henri Sohier
0.9	06/06/2024	Présentation à la reunion hebdomadaire	Henri Sohier
1.0	25/06/2024	Première version finalisée	Henri Sohier
2.0	27/09/2024	Prise en compte des reviews ainsi que des retours sur la version publique	Henri Sohier

A. Résumé

Un programme de R&D industrielle ambitieux en intelligence artificielle

Confiance.ai est un programme de recherche technologique français visant à permettre aux industriels l'intégration de l'IA de confiance dans leur systèmes critiques.

Pilier technologique du Grand Défi « Sécuriser, certifier et fiabiliser les systèmes fondés sur l'intelligence artificielle » lancé par l'État dans le cadre de France2030, Confiance.ai est né de l'ambition de faire de la France un des pays leader de l'industrialisation de l'IA de confiance.

Les travaux de Confiance.ai ont permis, par leur ampleur, de travailler sur de nombreuses problématiques de l'IA de confiance et de faire émerger des solutions logicielles et méthodologiques. Confiance.ai couvre un large périmètre, depuis les problématiques de qualité et de risque jusqu'aux problématiques de robustesse et d'embarquabilité.

Une nouvelle réglementation européenne

Le règlement européen sur l'intelligence artificielle, ou AI Act, est une "législation produit" de l'Union européenne visant à réguler les systèmes/produits d'IA commercialisés sur le marché européen, en fournissant aux opérateurs impliqués dans la chaîne de l'IA (développeurs/fournisseurs, importateurs, distributeurs, fabricants, utilisateurs) un cadre juridique visant à parer les risques sur la santé, la sécurité et les droits fondamentaux que peut apporter cette technologie.

Ce règlement repose sur une approche dite "par les risques" dépendant de l'usage et de la finalité prévue du système d'IA – et non de la technologie. Quatre niveaux de risque sont prévus par l'AI Act, en plus des risques liés à l'IA générative.

Le règlement prévoit des exigences claires, une évaluation de la conformité avant que le système d'IA soit mis en service ou commercialisé, ainsi qu'une application de la réglementation après sa mise sur le marché. Enfin, une structure de gouvernance aux niveaux européen et national est établie.

L'AI Act a été publié au journal officiel en juillet 2024.

Des points de vue qui se nourrissent

Les travaux de Confiance.ai sont ainsi rendus d'autant plus importants par la réglementation européenne qui nécessite des solutions technologiques pour être d'abord comprise, puis respectée. Le lien entre la réglementation et les travaux technologiques de Confiance.ai se fait en particulier au travers de la standardisation. La standardisation, par nature ouverte à tous, offre un cadre technique aligné sur la réglementation.

Ce document détaille le lien entre Confiance.ai et la réglementation européenne, lien qui a été progressivement renforcé par des actions transverses durant les quatre années du programme Confiance.ai..

B. Rappels sur l'AI Act

B.1 Documents clés

Cette section rassemble des documents clés en lien avec l'AI Act. Il peut s'agir de documents officiels qui, bien que publics, peuvent être difficiles à trouver. Il peut aussi s'agir de présentations ou de synthèses considérées comme utiles.

- Requête de standardisation de la Commission Européenne auprès du CEN-CENELEC [1]
- AI Act approuvé par le Conseil [2]
  - Chapitre III : Système d'IA à haut risque
  - Annexe I Sections A et B et Annexe III : Systèmes à haut risque
- Questions et réponses sur l'IA par la Commission Européenne [3]
- Page de présentation de l'AI Pact [4]
- Page de présentation de l'AI Office [5]
- Briefing sur les bacs à sable réglementaires par le Parlement Européen [6]
- Synthèses sur l'AI Act :
  - "Ten essential EU AI Act questions businesses need to know" par KPMG [7]
  - "All you need to know to understand and comply with the EU law on AI" par France Digitale et Wavestone [8]

B.2 Systèmes considérés

B.2.1 Catégorisation par niveaux de risques initiaux

L'approche basée sur les niveaux de risque impose des obligations à tous les acteurs de la chaîne de l'IA (développeurs/fournisseurs, importateurs, distributeurs, fabricants, utilisateurs). L'AI Act considère quatre niveaux de risque, en plus des risques systémiques de l'IA générative :

Risque minimal ou nul : La grande majorité des systèmes d'IA actuellement utilisés dans l'UE appartiennent à cette catégorie. La proposition permet l'utilisation libre des IA à risque minimal. Volontairement, les fournisseurs de ces systèmes peuvent choisir d'appliquer les exigences pour une IA de confiance et de se conformer aux codes de conduite volontaires (Art. 69 – Codes de conduite). Lorsque des systèmes d'IA conformes présentent un risque, l'opérateur concerné sera tenu de prendre des mesures pour s'assurer que le système ne présente plus ce risque, de retirer le système du marché, ou de rappeler le risque pendant une période raisonnable proportionnelle à la nature du risque (Art. 67 – Systèmes d'IA conformes présentant un risque). Par exemple : jeux vidéo activés par IA ou filtres anti-spam.

Risque limité : Les systèmes pour lesquels les utilisateurs doivent être conscients qu'ils interagissent avec une machine afin de pouvoir prendre une décision éclairée de continuer à utiliser le système ou de retirer le système des opérations. Ces systèmes doivent se conformer à des obligations spécifiques d'information et de transparence ; par exemple, les chatbots et les systèmes générant des deepfakes ou du contenu synthétique.

Systèmes d'IA à haut risque : Il s'agit des systèmes qui peuvent avoir un impact significatif sur la vie d'un utilisateur (Art. 6) et affecter la sécurité des personnes ou leurs droits fondamentaux. Il s'agit par exemple d'IA utilisée dans l'éducation, le recrutement, l'accès aux services publics et privés essentiels (huit types de systèmes sont ainsi définis dans l'Annexe III). Il peut aussi s'agir de systèmes d'IA utilisés comme composants de sécurité de produits physiques déjà couverts par des réglementations européennes comme les avions, dispositifs médicaux, machines, jouets, etc. (Annex I section A et section B).

Risque inacceptable : Le chapitre II en son article 5 dresse certaines pratiques qualifiées « d'interdites » pour des raisons évidentes d'atteintes sévère à la sécurité, les moyens de subsistances et les droits fondamentaux des citoyens européens. La mise sur le marché, l'entrée en service ou l'usage de système d'IA de la liste suivante sont interdites : notation sociale, reconnaissance faciale, dark-patterns et systèmes d'IA manipulateurs, les systèmes d'assistance vocale qui encouragent des comportements dangereux, ou les systèmes d'identification biométrique à distance en temps réel dans les espaces publics pour l'application de la loi.

Les développeurs/fournisseurs de systèmes d'IA à haut risque sont les plus touchés par les obligations de l'AI Act.

La figure suivante représente ces niveaux de risques.

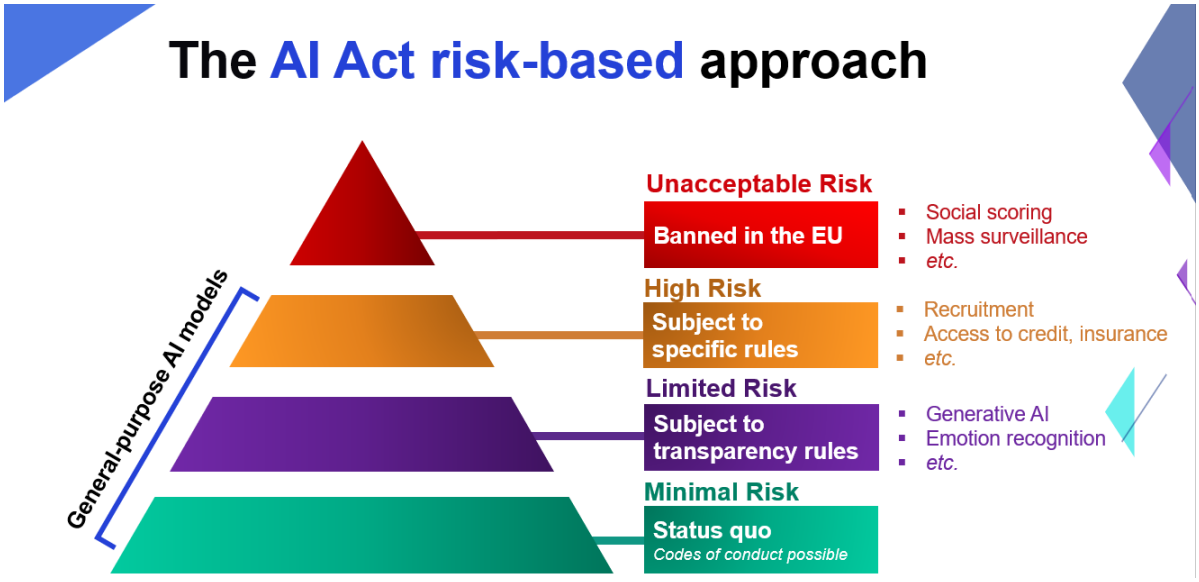


Figure 1 – L'approche basée sur les risques définie dans/par l'AI Act
Source : DIGITALEUROPE, 2024

B.2.2 IA à usage général (GPAI) et risques systémiques

Les fournisseurs de modèles à usage général (ou GPAI, pour « General Purpose Artificial Intelligence ») ont plusieurs obligations pour assurer le développement et le déploiement sûrs et responsables de leurs modèles. Selon l'article 53, ils doivent d'abord notamment :

- établir et tenir à jour une documentation technique, y compris des informations sur le processus d'entraînement et de test du modèle, à la fois pour le Bureau de l'IA et les fournisseurs de systèmes d'IA qui souhaitent intégrer le modèle dans leurs systèmes ;
- rendre publique un résumé du contenu utilisé pour l'entraînement du modèle.

Les fournisseurs peuvent s'appuyer sur des codes de conduite ou des normes harmonisées européennes pour démontrer leur conformité à ces obligations. Pour les modèles GPAI présentant des risques systémiques, les fournisseurs doivent notamment réaliser une évaluation du modèle et signaler et suivre les incidents graves.

Un modèle GPAI peut être classé comme posant un risque systémique s'il possède des capacités à fort impact (article 51(1), point (a)), évaluées à l'aide d'outils et de méthodologies techniques. Cela inclut des facteurs tels que :

- le nombre de paramètres,
- la qualité et la taille de l'ensemble de données,
- la quantité de calcul utilisée pour l'entraînement (>10^25 FLOPS),
- les modalités d'entrée et de sortie.

La portée du modèle et son accessibilité à un grand nombre d'utilisateurs, tels qu'au moins 10 000 utilisateurs professionnels enregistrés dans l'Union Européenne, peuvent également contribuer à sa classification en tant que risque systémique. De plus, la capacité du modèle à s'adapter à de nouvelles tâches, son niveau d'autonomie et son accès à des outils peuvent également augmenter son potentiel à poser des risques systémiques. La Commission devrait adopter des actes délégués pour modifier les seuils et les critères de référence si nécessaire pour refléter l'état de l'art. De plus, il est tout à fait possible qu'un modèle GPAI puisse également être classé en fonction d'une décision de l'AI Office, suite à une alerte qualifiée du panel scientifique indiquant qu'un modèle GPAI possède des capacités ou un impact importants.

B.2.3 Applications sectorielles

L'AI Act se veut d'application horizontale ("horizontal regulatory framework"), signifiant que les exigences du texte concernant les systèmes d'IA à haut risque doivent s'appliquer à divers secteurs. En ce qui concerne les systèmes d'IA destinés à être utilisés comme composants de sécurité de produits physiques déjà réglementés par l'UE et listés dans la section B de l'annexe 1 – ces systèmes d'IA sont d'ailleurs classés comme à haut risque –, comme c'est le cas pour les secteurs automobile et aéronautique, une application indirecte de l'AI Act va se faire afin que les produits liés restent principalement réglementés par leur cadre sectoriel. Les industries concernées resteront impactées par l'AI Act puisque les exigences techniques de la réglementation liées aux systèmes d'IA à haut risque seront adaptées aux spécificités du secteur en question. Cela se fera notamment via des actes délégués et/ou d'exécution de la Commission européenne (art. 102-110). Ces industries devront donc suivre de près les procédures d'élaboration des règles et standards de leur secteur ; par exemple, en ce qui concerne l'aviation civile, l'Agence européenne de la sécurité aérienne (EASA) et l'Organisation européenne pour l'équipement de l'aviation civile (EUROCAE) travaillent sur ces sujets.

Application automobile

La loi sur l'intelligence artificielle peut ainsi avoir un impact significatif sur le développement et le déploiement des véhicules autonomes (VA) ou des systèmes automobiles suivants : les Systèmes avancés d'aide à la conduite & Avertissement, les Systèmes d'évaluation du risque pour le conducteur, les Systèmes avancés de freinage d'urgence, le Régulateur de vitesse adaptatif, les Systèmes avancés de maintien de la trajectoire, l'Adaptation intelligente de la vitesse, la Surveillance de la somnolence et de la conscience et de vigilance, les Systèmes de recommandation pour la reconnaissance des panneaux de signalisation, l'Analyse émotionnelle du conducteur…

Il est par ailleurs à noter que l'Annexe 3, qui liste les cas d'usage des Systèmes à Haut Risques, mentionne le « trafic routier » qui évoque surtout des infrastructures routières telles que la communication, les signalisations, lumières, etc. L'Annexe 3 mentionne également les éléments destinés à être utilisés comme composants de sécurité ou qui sont eux-mêmes des produits nécessitant une évaluation de conformité par un tiers, ce qui peut affecter différents composants d'un VA. Finalement, l'art. 112 prévoyant une révision annuelle des cas d'usage des systèmes à haut risque pourrait conduire à mentionner directement les VA.

Dans l'ensemble, la classification à haut risque pour les systèmes d'IA utilisés dans les VA obligerait le secteur automobile à s'adapter à de nouvelles réglementations et à une nouvelle surveillance, tout en créant

de nouvelles opportunités d'innovation et d'investissement dans le développement de systèmes d'IA sûrs et fiables. Mais les acteurs de l'automobile ne sont pas unanimes sur la manière dont l'AI Act s'applique étant donné que les composants embarqués dans les véhicules d'une manière générale sont régulés à travers une approche sectorielle intitulés « type-approval » alias « certificat d'homologation ».

Application aéronautique

Le domaine aéronautique est fortement concerné par la technologie de Machine Learning, apportant des aides précieuses : Détection de pistes d'atterrissage, détection des dommages sur l'avion permettant de faciliter la maintenance, détection de la présence d'oiseaux aux alentours des pistes d'atterrissage, aide à la décision, etc.

Les spécificités de l'intelligence artificielle à l'égard des systèmes d'aviation civile à haut-risque, notamment pour l'aspect sûreté de fonctionnement (« safety »), ont conduit à l'évolution de la réglementation aéronautique hors IA (UE) 2018/1139 [9] à travers l'article 108 de la loi de l'intelligence artificielle.

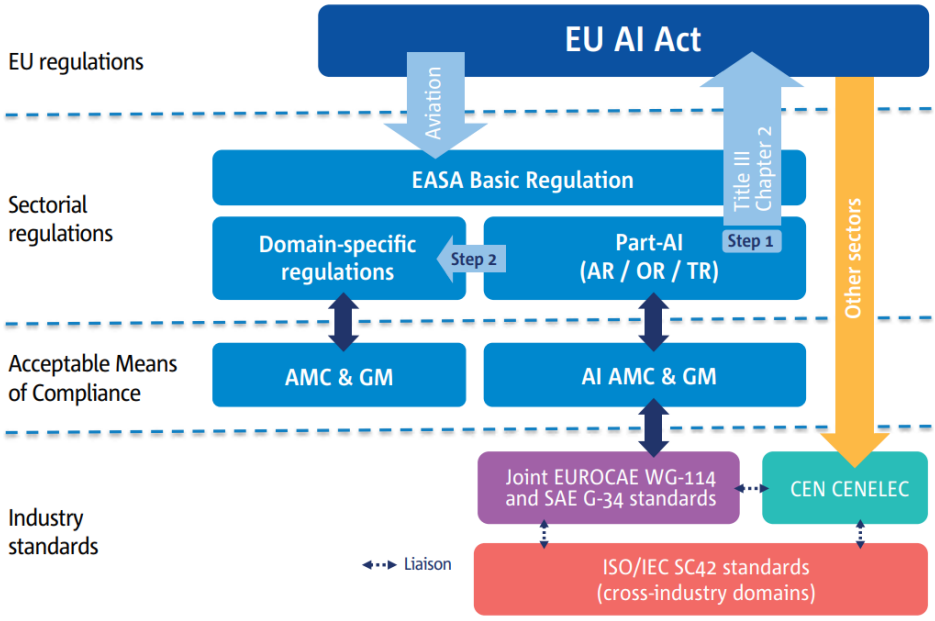


Figure 2 - Lien entre l'AI Act et la réglementation aéronautique
Source : EASA AI Roadmap [10]

Le document EASA Concept Paper [11] vise à fournir un guide aux industriels pour tous les domaines couverts par le règlement « EASA Basic Regulation / (EU) 2018 – 1139 » de l'aviation civile. Ce document est composé de 4 blocs : Analyse d'une IA de Confiance, Assurance IA, Facteurs Humains pour l'IA, Réduction des risques liés à l'aspect Safety.

Le WG EUROCAE 114 | SAE G-34 est un groupe de réflexion et de rédaction du nouveau standard L'ARP-6983 | ED-324 « Recommended Practice for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing ML » [12]. La rédaction de cette première version du standard est

limitée au « Non-adaptive Machine Learning in Supervised Mode ». Ce nouveau standard s'intéresse à l'intégration des composants ML dans les systèmes critiques. Il définit un ensemble de méthodes, de processus et des objectifs qui vont aider à la certification des systèmes à base de ML dans le domaine avionique d'une part, et dans le domaine du traffic aérien et des services de navigation aériennes (ATM/ANS) d'autre part. Ce document pourrait devenir une AMC (« Acceptable Mean of Compliance ») pour l'EASA.

B.3 Lien avec la standardisation

B.3.1 Requête de standardisation

L'AI Act étant une législation de l'UE appartenant à la "Nouvelle approche" ("New Legislative Framewok"), il s'accompagne étroitement d'une solution technique l'opérationnalisant. En effet, l'AI Act fixe des grandes lignes directrices et objectifs à atteindre pour assurer la santé, sécurité et respect des droits fondamentaux des personnes à l'égard des systèmes d'IA mis sur le marché européen, et se veut soutenu par une solution technique décrivant comment opérationnellement atteindre ces objectifs.

Dans le cadre de l'AI Act, la Commission européenne a mandaté en mai 2023 le CEN-CENELEC, via une requête de standardisation, pour publier des standards horizontaux, et ce en les développant lui-même ou en adaptant et adoptant ceux de l'ISO. Cette relation entre la réglementation et les standards est représentée dans la figure suivante.

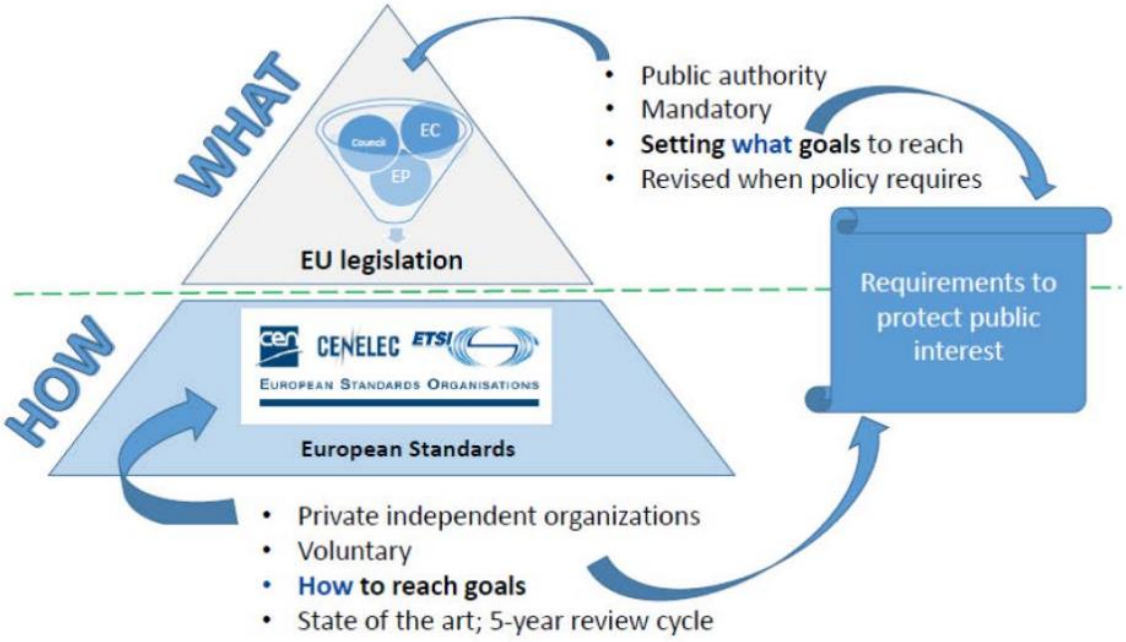


Figure 3 – Relation entre la réglementation et les standards

Cette première requête de standardisation (voir détails dans le tableau ci-dessous) se concentre, en 10 points, sur les exigences demandées pour les systèmes d'IA à haut risque. Des standards pour les IA à usage général/IA générative seront par la suite développés – un recours à des codes de bonnes pratiques émis par la Commission européenne devra entre temps être fait.

Tableau 1 – Thématiques de la requête de standardisation (travail Confiance.ai)

Thématique relevée dans la requête de standardisation	Articles de l'AI Act pouvant être considérés comme liés à la thématique
risk management system for AI systems	Article 9: Risk Management System
governance and quality of datasets used to build AI systems	Article 10: Data and Data Governance
record keeping through logging capabilities by AI systems	Article 12: Record-Keeping Article 19: Automatically Generated Logs
transparency and information provisions to the users of AI systems	Article 11: Technical Documentation Article 13: Transparency and Provision of Information to Deployers Article 18: Documentation Keeping Article 20: Corrective Actions and Duty of Information
human oversight of AI systems	Article 14: Human Oversight
accuracy specifications for AI systems	Article 15: Accuracy, Robustness and Cybersecurity
robustness specifications for AI systems	Article 15: Accuracy, Robustness and Cybersecurity
cybersecurity specifications for AI systems	Article 15: Accuracy, Robustness and Cybersecurity
quality management system for providers of AI systems, including post-market monitoring process	Article 17: Quality Management System
conformity assessment for AI systems	Article 43: Conformity Assessment

B.3.2 Travaux existants en standardisation

Le Tableau 2 représente des standards considérés pour harmonisation début 2023, et le Tableau 3 montre les liens possibles avec la requête de standardisation. L'attention du lecteur est attirée sur le fait que le Tableau 2 est relativement ancien et que la stratégie du CEN-CENELEC a sensiblement évolué. Beaucoup de standards ISO ne sont plus considérés comme harmonisable tel quel. L'annexe F offre de manière plus générale une revue des standards en lien avec la requête de standardisation.

Tableau 2 – Standards considérés pour harmonisation par le CEN-CENELEC JTC21 WG1, tel que présenté à la plénière du JTC21 du 16/17 Janvier 2023

ISO/IEC 22989	Artificial Intelligence concepts and terminology
ISO/IEC 23053	Framework for Artificial Intelligence (AI) system using Machine Learning
ISO/IEC 42001	AI management system
ISO/IEC 23894	AI Risk Management
ISO/IEC 5259-part 1	Data quality for analytics and machine learning (ML) - Overview, terminology, and examples
ISO/IEC 5259-part 2	Data quality for analytics and machine learning (ML) Data quality measures
ISO/IEC 5259-part 3	Data quality for analytics and machine learning (ML) Data quality management requirements and guidelines
ISO/IEC 5259-part 4	Data quality for analytics and machine learning (ML) Data quality process framework
ISO/IEC 27001:2013	Information security management systems
ISO/IEC 42006	Requirements on bodies performing audit and certification of AI management systems
CEN-CENELEC Risk	AI Risk catalogue and management
CEN-CENELEC Trustworthiness	AI trustworthiness characterisation

Tableau 3 – Lien possible entre les standards et la requête de standardisation (travail Confiance.ai)

Thématique relevée dans la requête de standardisation	Standards et normes
risk management system for AI systems	ISO 42001 ISO/IEC 23894 – Guidance on risk management ISO/IEC CD 42005 – AI system impact assessment
governance and quality of datasets used to build AI systems	ISO/IEC 8183 – Data life cycle framework ISO 42001 ISO IEC 5259 – Data quality ISO/IEC TR 24027 – Bias in AI systems and AI aided decision making
record keeping through logging capabilities by AI systems	
transparency and information provisions to the users of AI systems	ISO DIS 12792 – Transparency taxonomy of AI systems ISO 42001 ISO/IEC TR 24028 – Overview of trustworthiness in artificial intelligence
human oversight of AI systems	Not published yet: ISO/IEC AWI 42105 Guidance for human oversight of AI systems ISO/IEC TS 8200 Controllability of automated artificial intelligence systems
accuracy specifications for AI systems	
robustness specifications for AI systems	ISO/IEC TR 24029 – Artificial Intelligence (AI) — Assessment of the robustness of neural networks
cybersecurity specifications for AI systems	ISO/IEC TR 5469:2024 Functional safety and AI systems
quality management system for providers of AI systems, including post-market monitoring process	ISO/IEC TS 25058:2024 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SquaRE) — Guidance for quality evaluation of artificial intelligence (AI) systems
conformity assessment for AI systems	

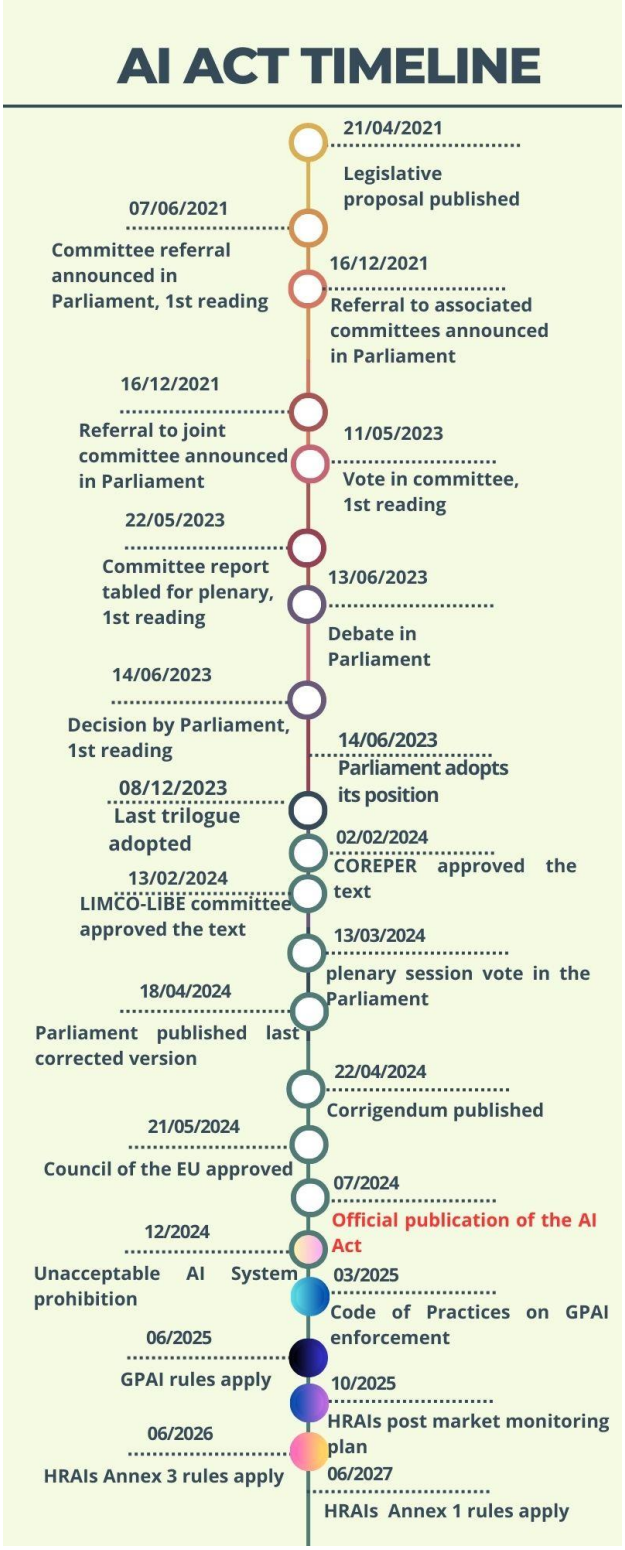
B.4 Chronologie

Le calendrier de l'AI Act débute le 8 avril 2019 avec la présentation par le groupe d'experts de haut niveau des lignes directrices en matière d'éthique pour une intelligence artificielle digne de confiance, mettant l'accent sur une IA éthique qui priorise le bien-être humain, la transparence et la responsabilité. En février 2020, le livre blanc de la Commission européenne sur l'IA a souligné la nécessité d'un cadre réglementaire pour équilibrer les avantages de l'IA et la protection des citoyens. Puis, le 17 juillet 2020, la liste finale d'évaluation de l'IA digne de confiance (ALTAI) a été introduite, offrant une liste de contrôle pratique pour la mise en œuvre des principes de l'IA digne de confiance. Ce calendrier reflète l'engagement de l'Europe à développer une IA légale, éthique et robuste, garantissant que l'IA renforce les capacités humaines et le bien-être de la société.

On peut ainsi noter ces premières dates clés :

- 2018 : Création du High Level Expert Group
- 2019 : Conclusions sur un plan coordonné pour l'IA
- 2019-2020 : Publications de rapports et documents sur l'IA de confiance par le « High Level Expert Group » (HLEG)
- 2020 : White Paper IA de la Commission Européenne
- Avril 2021 : introduction du texte par la Commission européenne (CE)

La progression du développement de l'AI Act est ensuite décrit dans la figure ci-contre. Il est possible d'y ajouter le souhait de publier les standards demandés au CEN-CENELEC en Avril 2025.



B.5 Structures pour l'implémentation de l'AI Act

B.5.1 AI Office

L'article 64 de l'Act instaure « le Bureau européen de l'IA qui sera le centre d'expertise en matière d'IA dans l'ensemble de l'UE. Il jouera un rôle clé dans la mise en œuvre de la législation sur l'IA, en particulier pour l'IA à usage général, favorisera le développement et l'utilisation d'une IA digne de confiance, ainsi que la coopération internationale. »

Le Bureau de l'IA est composé de plusieurs unités :

1) Réglementation et conformité, pour assurer l'application uniforme de la loi sur l'IA

2) l'unité de sécurité de l'IA, pour identifier les risques et les mesures d'atténuation associées pour les GPAI très performants

3) l'unité d'excellence en matière d'IA et de robotique, pour soutenir la R&D,

4) l'unité "AI for Societal Good", pour s'engager dans des projets d'IA bénéfiques,

5) l'unité "AI Innovation and Policy Coordination", pour superviser la stratégie de l'UE en matière d'IA, y compris les investissements, l'adoption de l'IA et les "bacs à sable" réglementaires.

Page de présentation de l'AI Office : https ://digital-strategy.ec.europa.eu/en/policies/ai-office

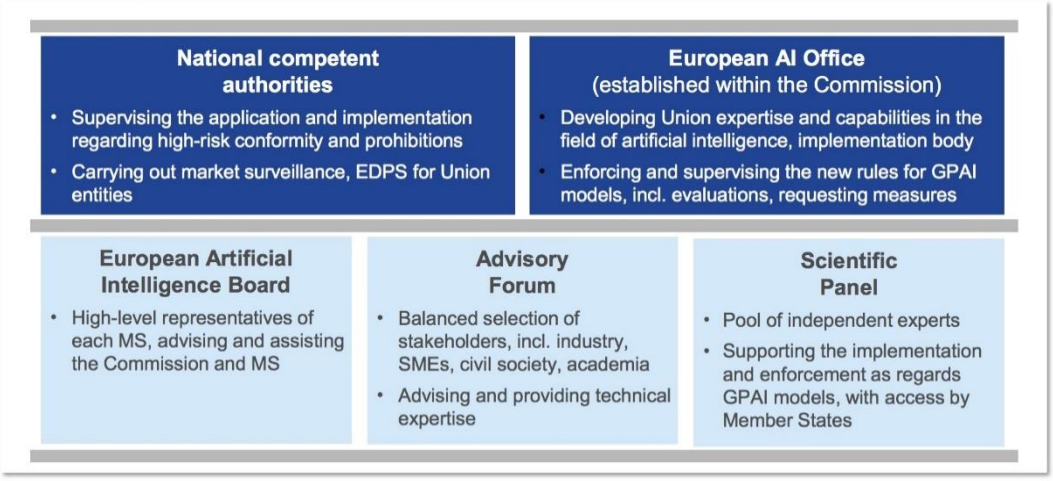


Figure 5 – Gouvernance pour l'application de l'AI Act
Source : AI Office, May 2024

B.5.2 Regulatory sandboxes

A travers l'article 57, « la loi sur l'intelligence artificielle envisage la mise en place de « bacs à sable réglementaires » coordonnés pour favoriser l'innovation en matière d'intelligence artificielle (IA) à travers l'UE. Un bac à sable réglementaire est un outil permettant aux entreprises d'explorer et d'expérimenter de nouveaux produits, services ou modèles d'affaires innovants sous la supervision d'un régulateur. Il offre aux innovateurs des incitations à tester leurs innovations dans un environnement contrôlé, permet aux régulateurs de mieux comprendre la technologie, et favorise le choix des consommateurs à long terme. Cependant, les bacs à sable réglementaires comportent également un risque d'utilisation abusive ou détournée, et nécessitent un cadre juridique approprié pour réussir. »

Briefing sur les sandboxes par le Parlement Européen de 06/2022 :
https ://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf

B.6 Ecosystème

B.6.1 AI Pact

L'AI Pact est une initiative lancée fin 2023 par la Commission européenne, basée sur le volontariat, pour anticiper la mise en œuvre de l'AI Act. Ses objectifs principaux sont de partager les bonnes pratiques, de développer des mesures d'implémentation et de faciliter des communications publiques pour parvenir à une compréhension commune du règlement. Les points clés de l'initiative incluent :

- L'encouragement à une conformité anticipée avec l'AI Act et promouvoir un dialogue ouvert avec les entreprises ;
- L'expression de l'intérêt industriel via un formulaire en ligne non contraignant, auquel plus de 500 entreprises ont répondu ;
- L'organisation d'ateliers pour favoriser cette initiative.
- L'engagement publique et spécifique des entreprises ayant signé l'AI Pact.

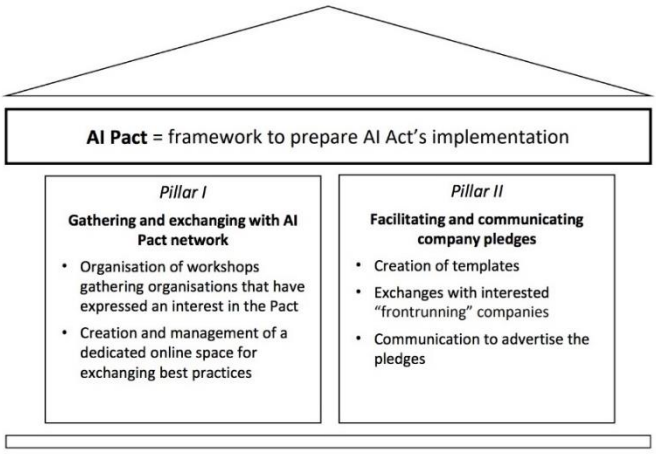


Figure 6 – Approche à deux pilliers de l'AI Pact (Source : AI Office, May 2024)

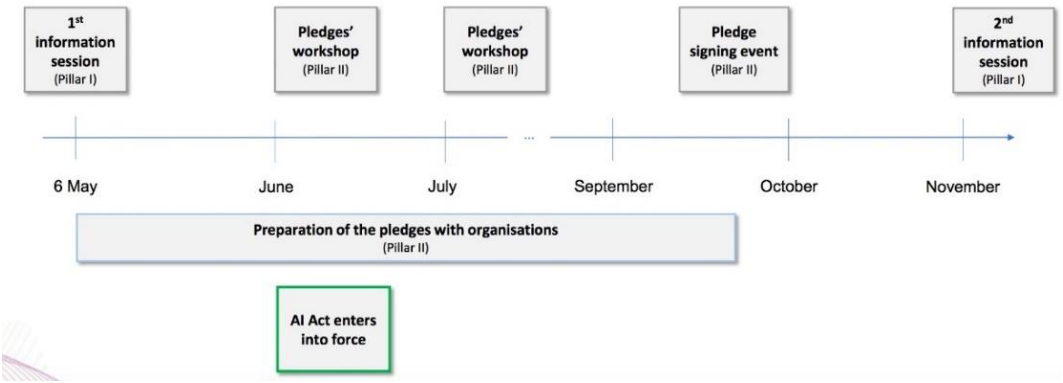


Figure 7 – Timeline souhaitée pour l'AI Pact (Source : AI Office, May 2024)

B.6.2 AI Trust Alliance (labélisation)

Afin d'anticiper la conformité avec l'AI Act, diverses initiatives de labels voient le jour. L'objectif de ces labels s'étend également, au-delà d'aspects réglementaires, à une démarche d'IA responsable. Les labels IA sont utilisés sur la base du volontariat et s'adressent à tous les systèmes d'IA (au-delà des systèmes à haut risque auxquels s'adresseront les standards harmonisés). Les labels détaillent la mise en œuvre d'exigences dans les thématiques mises en lumières par le règlement européen, mais ils peuvent également détailler cette mise en œuvre sur des thématiques connexes.

L'AI Trust Alliance est une initiative franco-allemande co-dirigée par l'institut allemand VDE, l'association française Positive AI, l'IEEE et Confiance.ai. Elle s'est construite sur l'idée d'intégrer les référentiels d'évaluations du VDE, de Positive AI et de l'IEEE de manière à produire un référentiel transverse européen.

B.6.3 DigitalEurope (trade association)

DigitalEurope est une association professionnelle qui représente l'industrie des technologies numériques en Europe en défendant les intérêts de ses membres et influençant les politiques et cadres réglementaires au sein de l'Union européenne.

L'association interagit avec les institutions de l'UE, les gouvernements nationaux et autres parties prenantes dans le but d'influencer les réglementations liées au numérique. Elle exprime les points de vue de l'industrie également à travers des prises de positions, répond aux consultations publiques, et publie des rapports.

De plus, DigitalEurope organise des événements, webinaires et conférences pour diffuser les connaissances et encourager les discussions entre les leaders de l'industrie et les décideurs politiques.

En ce qui concerne plus particulièrement l'AI Act, outre ces activités, DigitalEurope a permis à ses adhérents de suivre de près le processus législatif du règlement en leur donnant accès à des documents des institutions de l'UE non encore rendus publiques et en faisant intervenir des membres de ces institutions lors de sessions de l'association.

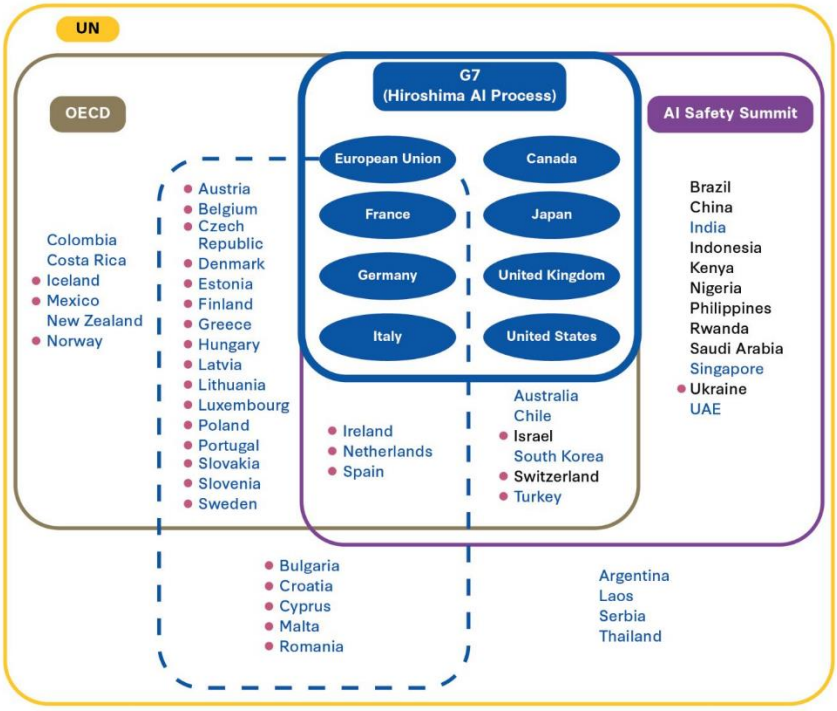
109 entreprises en sont actuellement membres (https://www.digitaleurope.org/corporate/).

B.7 Lien avec d'autres initiatives politiques

B.7.1 Processus d'Hiroshima

Le processus d'IA générative, connu sous le nom de processus d'Hiroshima du G7 sur l'intelligence artificielle (IA), a eu lieu en 2023 sous la présidence japonaise du G7. Cette initiative visait à aborder les principaux défis et opportunités liés à l'IA générative au sein des pays du G7. Par le biais de discussions, de l'établissement de priorités politiques et de la coopération internationale, le processus d'Hiroshima du G7 a cherché à faire progresser l'utilisation responsable et bénéfique des technologies de l'IA tout en en atténuant les risques.

Le processus d'Hiroshima du G7 sur l'IA identifie des défis tels que la prévention de l'utilisation abusive de l'IA générative pour les menaces et les risques de cybersécurité, tout en reconnaissant les opportunités d'innovation et de coopération. L'organisation a pour objectif de dresser un bilan, de hiérarchiser les politiques, de promouvoir la collaboration internationale, d'élaborer des codes de conduite et de soutenir la recherche afin de faire progresser l'IA générative de manière responsable au sein des pays du G7.



Notes: The European Union is considered a nonenumerated member of the G7. The countries shown in blue indicate those participating in the Hiroshima AI Process Friends Group (as of May 2, 2024). The nations with pink dots (•), plus the G7 members, are the members or observers of the Council of Europe, the host organization of the AI Treaty.

Source: Authors' own analysis

CSIS | WADHWANI CENTER FOR AI AND ADVANCED TECHNOLOGIES

Figure 8 – Paysage d'initiatives politiques internationales
Source : CSIS

B.7.2 L'US Executive Order

En octobre 2023, le président américain Joe Biden a adopté l'Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Ce décret vise à promouvoir le développement d'une IA responsable, en cherchant à équilibrer les avantages et les risques, notamment en établissant des lignes directrices en matière de sécurité et de protection de la vie privée, en promouvant l'innovation et en encourageant une coopération internationale.

B.7.3 Le NIST et l'Artificial Intelligence Risk Management Framework

Le National Institue of Standards and Technology (NIST) est une agence gouvernementale américaine dont la mission est de promouvoir l'innovation et la compétitivité industrielle des États-Unis en élaborant et en promouvant des normes dans des domaines comme la cybersécurité et les technologies de l'information. Le NIST a été chargé par l'Executive Order de diriger l'élaboration de lignes directrices clés en matière d'IA, et le cadre de maîtrise des risques liés à l'IA du NIST est mentionné à plusieurs reprises dans le décret.

Dans un premier temps, en janvier 2023, le NIST a publié un cadre de maîtrise des risques liés à l'intelligence artificielle (Artificial Intelligence Risk Management Framework, ou « AI RMF ») [9] pour aider les organisations à gérer les risques associés aux systèmes d'intelligence artificielle. Ce cadre n'est pas contraignant, mais fournit aux organisations des actions permettant de construire des systèmes d'IA dignes de confiance. Il se fonde sur 4 fonctions essentielles, à savoir : la gouvernance, la cartographie, la mesure et la gestion, qui visent à créer une approche globale de la maîtrise des risques tout au long du cycle de vie du système d'IA.

Figure 9 - Fonctions sur lesquelles se base la maîtrise du risque IA
Source : NIST

Plus récemment, en avril 2024, le NIST a publié le projet de cadre de maîtrise des risques spécifique aux IA génératives [10]. Ce nouveau document met en exergue de nouveaux risques liés à leur utilisation, tels que les biais, la désinformation et les "deepfakes".

C. Comprendre notre positionnement par rapport à l'AI Act

C.1 Différence par rapport au niveau réglementaire

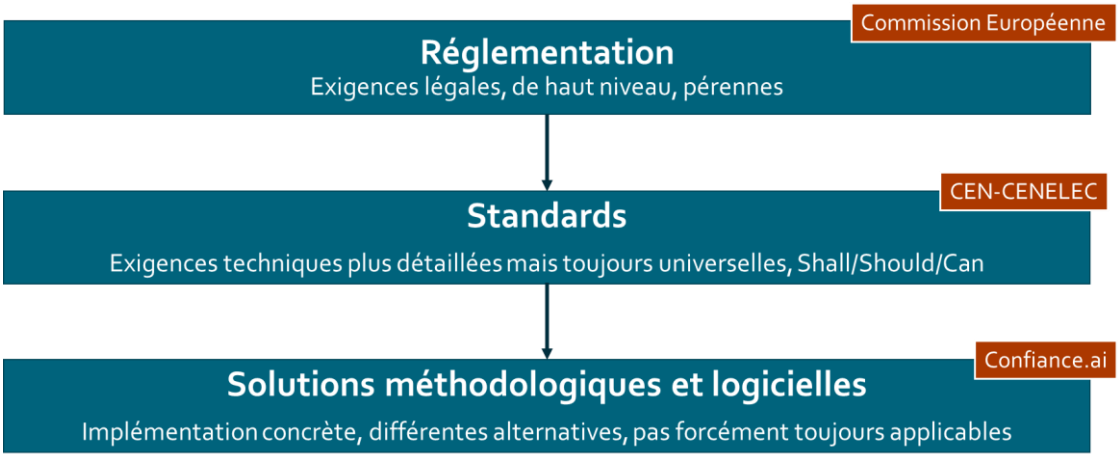


Figure 10 - Niveaux associés au niveau réglementaire

On peut analyser l'approche européenne sur l'intelligence artificielle de confiance comme constituée de trois niveaux.

La réglementation est le niveau supérieur : applicable à long terme, elle pose des exigences que les systèmes d'IA, notamment ceux qualifiés à haut risque, doivent satisfaire pour pouvoir être déployés au service des citoyens européens. Cela couvre notamment la transparence, traçabilité, robustesse, précision, contrôlabilité, etc.

Au niveau intermédiaire, les normes ou standards harmonisés, dont le développement a été confié à l'organisation CEN/CENELEC, préciseront la manière dont les exigences de haut niveau de la réglementation seront traduites en exigences concrètes pour les systèmes d'IA que les organisations devront mettre en œuvre et qui seront éventuellement vérifiées par des tierces parties (« notified bodies »).

Les normes harmonisées contiendront des exigences sur le processus de conception, de développement et de supervision des systèmes, ainsi que sur les produits « systèmes IA » eux-mêmes. Le troisième niveau est donc celui de la mise en œuvre concrète des exigences, et notamment des outils et méthodes permettant de les satisfaire. C'est à ce niveau que se situent les contributions du programme Confiance.ai, dont le sujet est justement la fourniture de méthodes et d'outils pour améliorer la confiance envers les systèmes d'IA pour des applications critiques. Les outils et méthodes produits par Confiance.ai visent d'abord les systèmes critiques, mais par extension peuvent être mis au service de tous systèmes d'IA quel que soit leur niveau de risque.

C.2 Systèmes considérés dans Confiance.ai

C.2.1 Systèmes à haut risque et systèmes critiques

L'AI Act met en avant la notion de système à haut risque, qu'elle associe aux usages pouvant typiquement impacter les droits ou la sécurité des consommateurs. Comme décrit en section B.2.1, il s'agit par exemple d'IA utilisée dans l'éducation ou le recrutement, mais aussi d'IA utilisée comme composant de sécurité.

Le Programme Confiance.ai s'est quant à lui construit sur la notion de « système critique », en la liant souvent à des questions industrielles de sécurité. L'accent est ainsi mis sur les systèmes « safety-critical ». Bien que proche de la notion de système à haut risque, la notion de système critique présente donc ici certaines différences. En particulier, l'expression « système critique » fait moins appel, dans Confiance.ai, à la notion de « droit » ou de « consommateur ». La notion de système critique peut toutefois avoir de nombreux sens en dehors de Confiance.ai. L'accent peut ainsi être potentiellement mis, dans d'autres contextes, sur les systèmes « business-critical », « mission-critical » or « security-critical ».

Tandis que le périmètre des systèmes à haut risque doit être défini de manière extrêmement précise dans l'AI Act, de manière à déterminer les exigences à respecter, la notion de système critique est considérée de manière plus flexible dans Confiance.ai (jusqu'à couvrir par exemple de l'opinion mining).

C.2.2 IA d'usage général

Comme décrit en section B.2.2, l'AI Act impose également des contraintes particulières à l'IA d'usage général. Ce sujet, qui s'est imposé relativement tardivement dans le développement de la réglementation, a été relativement peu couvert dans Confiance.ai. Il apparait ainsi peu dans le catalogue de Confiance.ai.

C.2.3 Produit, organisation, personnel

Les grands objectifs, notamment de sécurité, imposent des règles sur le produit, mais aussi sur la manière de faire le produit, sur l'organisation qui fait le produit, ou encore sur les personnes intervenant sur le produit. Ce cadre est typiquement appelé « système d'intérêt » en ingénierie système, ou « objet de conformité » en standardisation. Le système d'intérêt peut ainsi être le produit lui-même ou l'organisation qui fait le produit, par exemple.

Cette distinction importante n'est pas toujours facile à percevoir. L'AI Act est clairement associé à une approche produit, étant donné qu'il impose un marquage sur le produit. Toutefois, de nombreuses exigences seront davantage sur la manière de faire le produit, que sur les fonctionnalités ou les performances du produit lui-même. Ces typiquement le cas des exigences de gestion de qualité et de risque. C'est une approche classique pour les exigences trans-sectorielles (horizontales).

Le programme Confiance.ai offre des solutions de différentes natures : de nouvelles fonctionnalités intégrées aux systèmes d'IA, des gains de performance sur des systèmes d'IA donnés, des méthodes et outils pour améliorer le développement d'IA.

C.3 Lien avec les thématiques de la requête de standardisation

Les résultats recensés dans le catalogue de Confiance.ai ont été associés à une ou plusieurs thématiques de la requête de standardisation. Ce travail a aujourd'hui été réalisé pour 46 logiciels et 105 documents.

C.3.1 Contributions technologiques

Les apports technologiques de Confiance.ai à la réglementation européenne concernent principalement trois exigences, matérialisées dans trois des dix demandes de normes faites au CEN/CENELEC par la Commission Européenne :

- **la robustesse**, soit la capacité du système à réaliser la fonction prévue en présence d'entrées anormales ou inconnues. Confiance.ai a produit et testé une vingtaine de composants logiciels sur ce sujet.
- **L'exactitude** (accuracy) : mesure quantitative de l'ampleur de l'erreur de sortie du système d'IA. Une dizaine des composants testés et réalisés dans le cadre de Confiance.ai sont pertinents pour améliorer l'exactitude ou la précision des systèmes.
- **La qualité des données** : la mesure dans laquelle les données sont exemptes de défauts et possèdent les caractéristiques souhaitées pour l'application visée. Sur ce sujet également, Confiance.ai a produit et évalué une dizaine de composants ainsi qu'une plate-forme spécifique rassemblant diverses approches permettant d'améliorer la qualité des données d'entrée des systèmes d'apprentissage automatique.

Ces outils sont référencés dans le catalogue produit par confiance.ai, et plusieurs d'entre eux en *open source*, y sont disponibles en téléchargement.

Les contributions technologiques de Confiance.ai ne s'arrêtent pas à ces trois propriétés : nous avons également des contributions sur **d'autres exigences de la réglementation**, traduites dans des demandes de normes européennes harmonisées : sur **l'explicabilité** (correspondant aux besoins de transparence et de compréhension des systèmes d'IA par les humains) ; sur la **cybersécurité** des systèmes d'IA, notamment par le tatouage des productions des systèmes d'IA – images ou autres contenus; et plus généralement sur la **maîtrise des risques** et l'**analyse de conformité**. Là aussi, tous les composants logiciels sont référencés dans le catalogue en ligne et certains sont téléchargeables.

C.3.2 Contributions méthodologiques

Les contributions méthodologiques de Confiance.ai portent sur **le processus de développement d'un système d'IA**, qui va de la spécification initiale et la conception jusqu'à la mise en service et la supervision de son fonctionnement, y compris dans des systèmes embarqués. Ces contributions sont multiples :

- Une **taxonomie** des concepts et termes utilisés pour l'IA de confiance ;
- Une **documentation complète du processus**, comprenant une modélisation des activités et des rôles, avec les éléments permettant aux ingénieries des entreprises de le mettre en œuvre ;
- Un premier développement d'une **ontologie de l'IA de confiance**, reliant les principaux concepts du processus et de la taxinomie ;
- Et un « **corpus de connaissances** » (*Body of Knowledge*) qui regroupe l'ensemble de ces éléments et les rend accessibles sur le site web du même nom.

Si l'utilisation de ces outils méthodologiques n'est pas à elle seule une garantie de la conformité du système d'IA à la réglementation, elle peut en constituer un élément de justification, considéré comme faisant partie de l'état de l'art par les tierces parties (*notified bodies*) en charge de la vérification.

C.4 Lien avec les exigences des standards

Comme décrit en section B.3.2, le CEN-CLC/JTC 21 élabore actuellement les normes européennes qui seront compatibles avec la réglementation sur l'intelligence artificielle. Ces normes définiront des exigences à respecter et il est important d'identifier le lien entre les résultats de Confiance.ai et ces exigences. Il s'agit de liens plus fins qu'en section C.3.

Une norme européenne sera par exemple nécessaire sur le sujet de la qualité. Dans l'attente de la publication d'une telle norme, il est fréquent de se référer à l'ISO 42001 qui traite également ce sujet.

L'ISO 42001 est une norme établissant des exigences pour aider les organisations à mettre en place, à maintenir et à améliorer leur système de management. Le standard couvre l'établissement, la mise en œuvre, la maintenance et l'amélioration de la gestion d'un système d'IA. Il promeut une approche proactive et systématique pour identifier, évaluer et gérer les risques tout au long du cycle de vie des systèmes d'IA. Il est à noter que l'ISO 42001 considère des situations de différentes natures, potentiellement positives et négatives. A l'inverse, l'AI Act interprète le risque de manière plus spécifique.

Les exigences de l'ISO 42001 sur le sujet « Documentation and implementation of data management processes » peuvent par exemple être associées à ces résultats méthodologiques de Confiance:

- Methodology for Dataset Development (https://invenio.apps.confianceai-public.irtsysx.fr/records/t88d9-a9114)
- State of the Art on Data Engineering (https://catalog.confiance.ai/records/j6y20-q8p47)

De manière similaire, l'exigence de l'ISO 42001 "Implement mechanism to achieve human oversigh (Human reviewers; Monitor the performance of AI system; Reporting on outputs)" peut par exemple être associée aux résultats suivants:

- MAIAT (https://catalog.confiance.ai/records/eetp7-gz437)
- Formal XAI using Pyxai for anomaly detection (https://catalog.confiance.ai/records/a1bwm-w9906)
- Confidence score - Particul (https://catalog.confiance.ai/records/n6knc-fy315)

Le tableau suivant décrit l'ensemble des liens identifiés. Les résultats de Confiance.ai pourront ainsi être associés aux exigences des futures normes européennes lorsque celles-ci seront publiées.

Tableau 4 - Lien entre Confiance.ai et les exigences de standards

Standardization request	Standards qui pourraient représenter une partie de la réponse à la requête	Exemple d'exigences du standard	Outils Confiance.IA
Risk management	ISO 42001	Document a policy for the development or use of AI systems.	MAPIE (irtsysx.fr)
		Define and put in place a process to report concerns about the organization's role with respect to an AI system throughout its life cycle.	Documentation • Risk Management for AI-Based System • State of the Art on Risk Management of AI-Based Systems • Methodological Guidelines for Model ODD Characterization • Contribution on ODD for standardization • Methodological Guideline for ODD • Use Case Applications of Model ODD Characterization • ACAS-Xu implementation & Certification • End-to-end approach for engineering trusted AI-based systems • Methodological Guidelines for Rule-based Monitoring • Methodological Guideline for Multi-Scale Online Monitoring Framework • Use Case Applications of Rule-based Online Monitoring • Apply OOD detection to selected use cases • Use Case Applications of Monitoring Verification Tools Software • Online Monitoring Libraries • MOODD • Conformal Anomaly Detection • Topological Data Analysis for Anomaly Detection (TDAAD)

			• Machine learning TestBench for embedded implementation
Governance and quality of datasets used to build AI systems	ISO 42001	Documentation and implementation of data management processes	• Methodology for Dataset Development • Analysis and characterization of multicore and GPU implementations of NN (check) • State of the Art on Data Engineering •
		Documentation on data acquisition	• State of the Art on Synthetic Data Production and Integration • Guidelines on synthetic data usage in Machine Learning context
		Set requirements for data quality	• Characterization and Metrics for data coverage and quality
		Process for recording data provenance	
		Criteria for selecting data	• State of the Art on Data and Dataset Specification for AI
		Verification/Preparation of data	• Methodological Guideline for Time Series Anomaly Detection • State of the Art on Dataset Distances and Splitting • Leveraging Unlabeled Data in Active Learning for Detection • Guidelines on synthetic data usage in Machine Learning context • Conformal Anomaly Detection (anomaly detection – to check) • Topological Data Analysis for Anomaly Detection (TDAAD) (anomaly detection – to check) • Sparsity Based Anomaly Detection Framework (anomaly detection – to check) • Influenciae (anomaly detection – to check)
		Bias detection and bias treatment ⇨ ISO sur les biais	•

Specific to models?		Training phase	• State of the Art on Exploiting Unlabelled or Partially Labelled Dataset • State of the Art on Incremental Learning for Object Detection and Semantic segmentation • Methodological Guideline for Smart Data Management in an Iterative Context (todo: check again this one)
Record keeping through logging capabilities by AI systems	ISO 42001	Establish when the record keeping should be enabled	
Transparency and information provisions to the users of AI systems	ISO 42001	Documentation on data provenance, how data is used…	• Cartography of the moral situation in AI systems (could be also in SR1) • Unsupervised Anomaly Detection and Explainability tools for Time Series
		Provide clear information to users/customers: • Technical documentation • Risks related to the system • Results of impact assessment • Logs or system records	• Explainability Platform Kaa • report on Counterfactual-based-Metrics for the evaluation of an Image Classifier • Methodological Guideline for Explainability • Characterization of the notion of trust • Application of Similar Examples to Renault Welding Inspection • Formal XAI using Pyxai for anomaly detection • Benchmarking Environment Specification • Explicability benchmark • Influenciae • AIX360
Human oversight of AI systems		Implement mechanism to achieve human oversight, e.g.: • Human reviewers; • Monitor the performance of AI system; • Reporting on outputs;	• Leveraging Unlabeled Data in Active Learning for Detection (active learning: check if ok) • Semial (active learning) • Application of Similar Examples to Renault Welding Inspection • Formal XAI using Pyxai for anomaly detection

			• End-to-end approach for engineering trusted AI-based systems • MAIAT • Confidence score - Particul • Attribution-based confidence score - ReIdentification • Confidence score - True Class Probability (TCP)
Accuracy specifications for AI		Monitor accuracy of the AI system outputs	• Methodological Guideline for Time Series Anomaly Detection • Attribution-based confidence score - Classification (irtsysx.fr) • Randomized Smoothing for Classification (irtsysx.fr) • Topological Data Analysis for Anomaly Detection (TDAAD) (irtsysx.fr) • MAPIE (irtsysx.fr) • Attribution-based confidence score - Detection (irtsysx.fr) • Attribution-based confidence score - ReIdentification (confiance.ai) • Confidence score - True Class Probability (TCP) (confiance.ai)
Robustness specifications for AI systems			• Methodological Guideline and Evaluation tools for robust artificial intelligence (confiance.ai)
Cybersecurity specifications for AI systems		Communication of incidents, e.g. data breach	• Confidence score - True Class Probability (TCP) (confiance.ai) • Methodological Guideline and Evaluation tools for robust artificial intelligence (confiance.ai) • ML watermarking (irtsysx.fr)
quality management system for providers of AI systems, including post-market monitoring process			• Methodology to characterize and assess Trust for AI-based safety critical system (irtsysx.fr) • Taxonomy (irtsysx.fr)

Conformity assessment for AI systems		In case of non-conformity, implement a process for corrective actions: • React to non-conformity • Evaluate the need for action to eliminate the causes • Implement any action needed • Review effectiveness • Make changes to AI system management if needed.	• Regulation constraints compliance (Implementation Process Assurance) • Confiance Process Compliance With SAE AS6983 • Methodological Guideline for Assurance Cases Evaluation • Certification Process Implementation • Conformity With IEEE 7000 • State of the Art on Assurance Argumentation for the Efficient Qualification of AI-based Systems • State of the Art on Verification and Validation Strategies for AI-Based Systems • Certification of ML implementation
		Document all the steps.	

D. Conclusion

L'AI Act représente une avancée réglementaire majeure pour l'Union européenne. Visant à encadrer les systèmes d'intelligence artificielle en fonction de leurs niveaux de risque, elle impactera sensiblement l'industrie française.

L'AI Act impose des obligations strictes sur divers types de systèmes, en particulier ceux classés comme à haut risque. Ces obligations incluent une évaluation rigoureuse avant la mise en service ou la commercialisation des systèmes d'IA, ainsi qu'une surveillance continue après leur déploiement. Les détails des exigences seront précisés dans des standards harmonisés qui sont actuellement en cours de développement.

Les grandes thématiques abordées par l'AI Act, telles que la gestion des risques, la gouvernance des données, la transparence, la supervision humaine, et la robustesse des systèmes, correspondent largement aux axes de travail de Confiance.ai. Par conséquent, les outils développés dans le cadre de Confiance.ai s'avèrent particulièrement pertinents pour aider les entreprises à satisfaire aux exigences de la réglementation et des normes associées. Cependant, la manière dont ces outils faciliteront concrètement la conformité à l'AI Act est complexe à évaluer (a fortiori en l'absence des standards harmonisés).

E. Bibliography

[1] European Commission, "COMMISSION IMPLEMENTING DECISION on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence," 2023. [Online]. Available: https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en.

[2] EUROPEAN UNION, "REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008," 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:PE_24_2024_REV_1&qid=1719221715132.

[3] Commission européenne, "Intelligence artificielle - Questions et réponses," 2023. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/fr/QANDA_21_1683.

[4] European Commission, "AI Pact," [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/ai-pact.

[5] European Commission, "European AI Office," [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/ai-office.

[6] European Parliament, "Artificial intelligence act and regulatory sandboxes," [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf.

[7] KPMG, "Ten essential EU AI Act questions businesses need to know," 2024. [Online]. Available: https://assets.kpmg.com/content/dam/kpmg/ie/pdf/2024/01/ie-eu-artificial-intelligence-act.pdf.

[8] France Digitale and Wavestone, "All you need to know to understand and comply with the EU law on AI," 2024. [Online]. Available: https://media.francedigitale.org/app/uploads/prod/2024/02/01162803/Compliance-AI-Act-Feb-24.pdf.

[9] European Commission, "Regulation (EU) 2018/1139 of the European Parliament and of the council, on common rules in the field of Civil Aviation and establishing a European Union Aviation Safety Agency," 2018.

[10] EASA, "Artificial Intelligence Roadmap 2.0 - Human-centric approach to AI in aviation," 2023.

[11] EASA, "EASA Concept paper - Guidance for Level 1 & 2 machine Learning applications," 2024.

[12] EUROCAE / SAE, "EUROCAE ED-324 / SAE ARP6983 Draft5B - Recommended Practice for Development and Certification / Approval of Aeronautical Safety-Related Products Implementing ML," 2024.

[13] NIST, "Artificial Intelligence Risk Management Framework," 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[14] NIST, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," 2024. [Online]. Available: https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf.

F. Annexe : Standards

Requirements / SDO	Data and data governance	Risk management system	Technical data and Record keeping	Transparency and information to users	Human oversight	Accuracy, robustness, and cybersecurity	Quality management system
ISO and ISO/IEC JTC1	**ISO/IEC 25024;** ISO/IEC 5259; ISO/IEC 24668;	ISO/IEC 4213; ISO/IEC 25059; ISO/IEC 24029-2	ISO/IEC 5338; ISO/IEC 5469; ISO/IEC 24368; ISO/IEC 24372; ISO/IEC 24668	ISO/IEC 24027; **ISO/IEC 24028;** ISO/IEC 5338; ISO/IEC 24368; ISO/IEC 24372; ISO/IEC 24668; ISO/IEC 4213		ISO/IEC 24027; **ISO/IEC 24028;** ISO/IEC 24029; ISO/IEC 5469	ISO/IEC 23894; ISO/IEC 38507; ISO/IEC 42001; ISO/IEC 25059
IEEE	**ECPAIS Bias;** IEEE P7002; IEEE P7003; IEEE P7004; IEEE P7005; IEEE P7006; IEEE P7009; IEEE P2801; IEEE P2807; IEEE P2863	IEEE P7009; IEEE P2807; IEEE P2846	**ECPAIS Transparency;** IEEE P7000; IEEE P7001; IEEE P7006; IEEE P2801; IEEE P2802; IEEE P2807; IEEE P2863; IEEE P3333.1.3	**ECPAIS Bias; ECPAIS Transparency; ECPAIS Accountability** ; IEEE P7000; IEEE P7001; IEEE P7003; IEEE P7004; IEEE P7005; IEEE P7007; IEEE P7008; IEEE P7009; IEEE P7011; IEEE P7012; IEEE P7014; IEEE P2863; IEEE P3652.1	**ECPAIS Accountability; ECPAIS Transparency;** IEEE P7000; IEEE P7006; **IEEE 7010;** IEEE P7014; IEEE P2863	**ECPAIS Transparency;** IEEE P7007; IEEE P7009; IEEE P7011; IEEE P7012; IEEE P2802; IEEE P2807; IEEE P2846; IEEE P2863; IEEE P3333.1.3	IEEE 2801; IEEE P2863; IEEE P7000
ETSI	DES/eHEALTH-008 ; **GR CIM 007 ; GS CIM 009; ENI GS 001**; GR NFV-IFA 041; DGR SAI 002; TR 103 674; TR 103 675; TS 103 327; TS 103 194; TS 103 195.2,	**GS ARF 003 ; GR CIM 007 ; ENI GS 005**; GR NFV-IFA 041; DGS SAI 003; EG 203 341;	DES/eHEALTH-008 ; **ENI GS 005 ;** DGR SAI 002, **SAREF Ontologies; GR CIM 007; GS CIM 009**	DES/eHEALTH-008 ; **GS CIM 009** ; DGR SAI 002; **SAREF Ontology**	DES/eHEALTH-008 ; DGR SAI 005	**GS ARF 003 ; GR CIM 007 ; ENI GS 001; ENI GR 007** ; DGR SAI 001; DGR SAI 002; DGS SAI 003; GR SAI 004; **GS ZSM 002**; TR 103 674; TR 103 675; TS 103 327;	
	SAREF Ontologies	**TS 103 194; TS 103 195.2;** TR 103 821;				**GS 102 181, GS 102 182**	
ITU-T	**ITU-T Y.3170**; ITU-T Y.MecTa-ML ; **ITU-T Y.3531 ; ITU-T Y.3172 ;** ITU-T H.CUAV-AIF ;ITU-T F.VS-AIMC ; **ITU-T Y.4470 ; Y.Supp.63 to ITU-T Y.4000 series**	ITU-T Y.qos-ml-arc ; **ITU-T Y.3172 ;** ITU-T H.CUAV-AIF ; ITU-T F.VS-AIMC ; **ITU-T Y.4470**		**ITU-T Y.4470 ;**		**ITU-T Y.3170;** ITU-T Y.qos-ml-arc; ITU-T Y.MecTa-ML ; **ITU-T Y.3531 ; ITU-T Y.3172 ;** ITU-T H.CUAV-AIF ; ITU-T F.VS-AIMC ; **ITU-T Y.4470**	

Table 1. Overall representation of mapped standards (already published standards are in bold)[1]

[1] Stefano NATIVI, Sarah De Nigris, AI Standardisation Landscape: state of play and link to the EC proposal for an AI regulatory framework, EUR 30772 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-40325-8, doi:10.2760/376602, JRC125952

Title : Position de Confiance.ai par rapport à l'AI Act

Keywords : Réglementation, Conformité, Standards, Robustesse, Explicabilité

L'AI Act est une importante avancée réglementaire européenne visant à encadrer les systèmes d'IA selon leur niveau de risque. Elle impose des obligations strictes, notamment pour les systèmes à haut risque. Les exigences précises seront définies dans des standards harmonisés en cours d'élaboration. Les thèmes principaux de l'AI Act (gestion de la qualité et des risques, gouvernance des données, supervision humaine ou robustesse) sont alignés avec les axes développés dans Confiance.ai. Cependant, le rôle exact des outils de Confiance.ai dans la conformité est complexe à évaluer.

Our partners