

## Décryptage

# INTELLIGENCE ARTIFICIELLE DANS UNE MACHINE : QUEL IMPACT SUR SON FONCTIONNEMENT ?

De nombreux domaines commencent à intégrer de l'intelligence artificielle (IA). Les systèmes d'IA rendent des services dans la vie quotidienne (à domicile, au travail), comme dans beaucoup de métiers. Ils améliorent désormais le fonctionnement des machines et ces dernières donnent aux IA la possibilité d'interagir avec le monde physique. Ce changement interpelle le préventeur sur la fiabilité de ces systèmes et sur les risques en matière de santé et de sécurité qu'ils pourraient représenter. Alimenté par la fiction et par la médiatisation de recherches en IA, ce questionnement devient prégnant. Qu'en est-il réellement ? Cet article propose d'apporter des éclairages sur la nature de l'IA, sur son implantation dans les machines et son impact sur leur fonctionnement.

---

**ARTIFICIAL INTELLIGENCE IN MACHINERY: HOW DOES IT IMPACT MACHINE OPERATION?** – *Many fields have begun to integrate artificial intelligence (AI). AI systems perform services in everyday life (smartphones and other smart devices, etc.) and in many jobs. They now improve the operation of machines and these machines give AI the possibility to interact with the physical world. This change has alerted OSH practitioners about the reliability of these systems and the health and safety risks they may pose. Fuelled by fiction and media coverage of AI research, this topic has taken on great significance. What is really at stake though? This article sheds light on the nature of AI, its use in machines and its impact on their operation.*

---

MICHAËL  
SARREY  
INRS,  
département  
Ingénierie des  
équipements  
de travail

---

### Des machines autonomes

Il est tentant d'assimiler les machines munies d'une intelligence artificielle (IA) aux machines autonomes. En effet, la notion d'autonomie renvoie à la « *faculté d'agir librement, indépendamment*<sup>1</sup> ». Cependant, pour intuitive qu'elle soit, cette relation n'est ni directe, ni fréquente. Ainsi, deux questions complémentaires se posent.

### Les machines dotées d'IA sont-elles autonomes ?

Le rapport technique ISO 29119-11 (*Cf. Pour en savoir plus*), traitant du test des systèmes logiciels intégrant de l'IA, définit les systèmes autonomes comme étant « *capables de travailler sans*

*intervention humaine pendant des périodes prolongées* ». Il n'est pas fait référence ici à la liberté d'action de la machine intégrant de tels systèmes, mais bien à sa capacité à travailler seule. Dès lors, de nombreuses machines automatiques, certaines très anciennes, dotées ou non d'IA, peuvent être qualifiées de machines autonomes.

Selon cette définition, une IA qui accroîtrait la robustesse de fonctionnement de la machine, comme la capacité à s'adapter à un environnement changeant, contribuerait à la rendre autonome en réduisant la fréquence des interventions humaines nécessaires à son ajustement. En revanche, si une IA apporte une capacité de prise



ENCADRÉ 1

**LES TECHNIQUES D'APPRENTISSAGE AUTOMATIQUE**

**Apprentissage supervisé :** Cette technique d'apprentissage se base sur un jeu d'exemples labellisés (c'est-à-dire, dont le résultat est identifié). Elle est principalement utilisée pour la classification.

**Apprentissage non supervisé :** Désigne une technique de mise en évidence de caractéristiques communes à plusieurs éléments d'un jeu de données non labellisées. Cette technique est utilisée notamment dans la recherche de motifs annonceurs d'évènements (par exemple, pour de la maintenance prédictive).

**Apprentissage par renforcement :** Il s'agit d'une méthode d'apprentissage itérative où le programme reçoit une « récompense » (ou une « punition ») en fonction de l'évaluation de son action. En cherchant à maximiser ses récompenses, le programme optimisera son action après chaque itération.

de décision autonome à une machine, par exemple de s'arrêter pour préserver la sécurité d'un opérateur, la machine en question n'en deviendrait pas, pour autant, autonome. L'utilisation d'IA dans une machine n'entraîne pas nécessairement son autonomie. Est-ce réciproque ?

**Les machines autonomes intègrent-elles de l'IA ?**

Les AGV (en anglais : *automatic guided vehicles*) sont des chariots de transport de biens autonomes et guidés. Autonomes, parce qu'ils déclenchent leur mission sans ordre humain explicite (les missions sont appelées par les équipements de l'atelier ou par le système de gestion de production) et guidés, parce qu'ils suivent une piste pré-établie. Les AGV sont des machines autonomes, mais aujourd'hui, en général, ces véhicules ne disposent pas d'IA dans leur système de commande. Ainsi, pour les machines, la question de l'autonomie n'est pas nécessairement liée à l'implantation d'une IA dans leur système de commande.

**Des machines apprenantes**

Mais qu'est-ce que l'IA si elle ne permet pas de rendre les machines plus autonomes ? La question de la définition de l'IA reste aujourd'hui sans réponse unique.

Dans le cadre des machines, l'automatique a apporté d'énormes capacités aux machines; l'autonomie (comme vu précédemment), la prise de décision logique, le séquençement de tâches complexes ou encore la perception de leur environnement. Cependant, pour complexes qu'elles soient, ces fonctions automatiques ont toutes en commun d'avoir été spécifiées, modélisées, implémentées et testées par des automaticiens humains. Ces fonctions permettent la conception

de machines automatiques qu'on qualifiera de « conventionnelles ».

Durant la décennie 2010, l'apprentissage automatique s'est développé grâce à la technologie des réseaux de neurones artificiels et à l'accroissement de la puissance de calcul des ordinateurs. Cette technologie a donné aux programmes d'ordinateurs la capacité d'apprendre. Cet apprentissage est réalisé par des algorithmes, sur la base d'une multitude d'exemples et non pas par la mémorisation et l'application de règles préétablies.

Ainsi, une machine qui possède la faculté d'apprendre ses actions par l'observation et la synthèse d'exemples de tâches sera qualifiée de « *machine intégrant une IA* ». Il s'agit d'un réel changement de paradigme dans la manière de concevoir le système de commande d'une machine. Le programme de la machine n'est alors pas écrit, ou même décrit, par un automaticien, mais il est déduit de l'observation des exemples qui lui sont soumis par un algorithme d'apprentissage (Cf. Encadré 1).

**Les méthodes de programmation et leurs conséquences sur les machines**

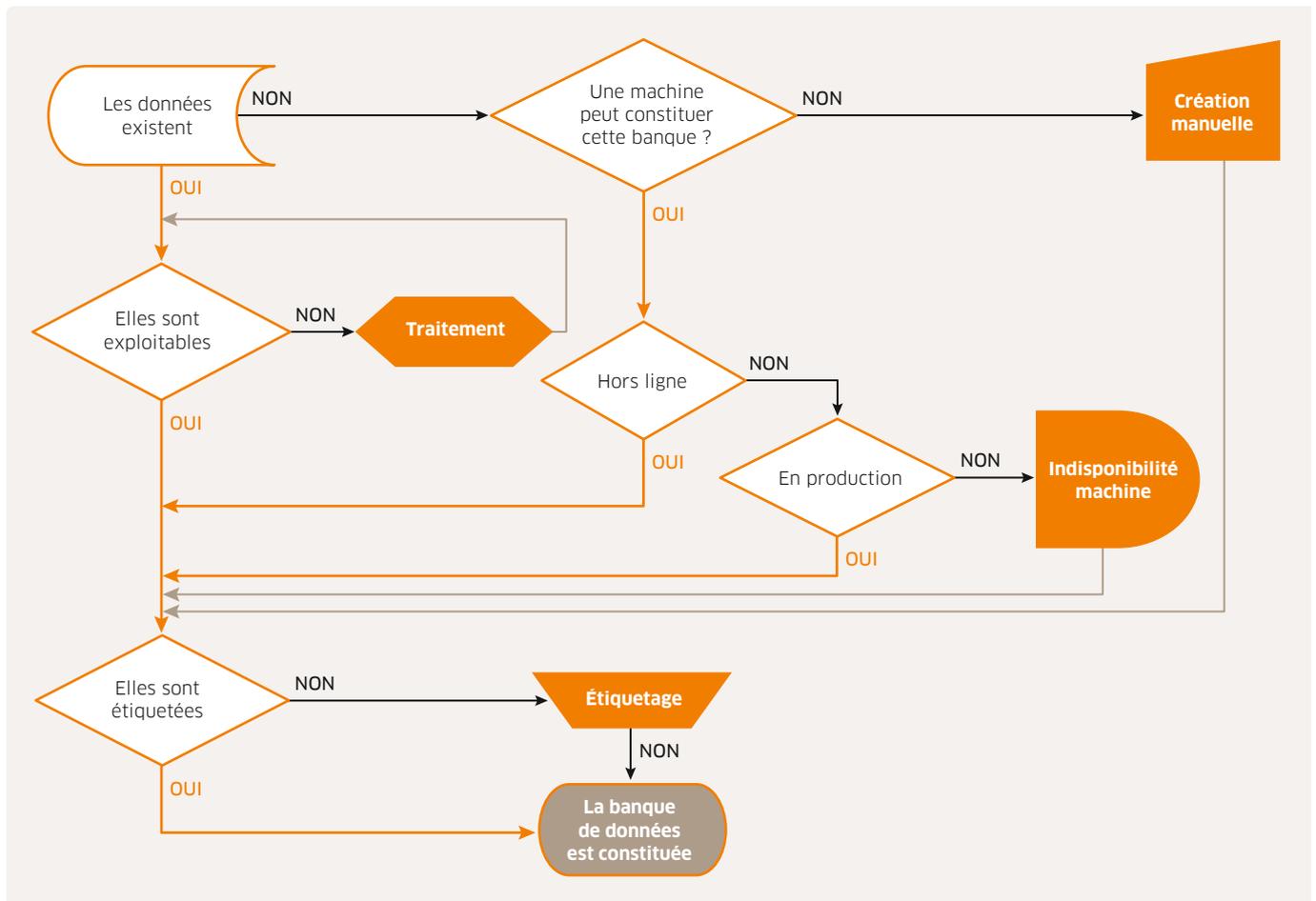
Deux voies permettent donc de programmer des machines automatiques : la voie analytique, où un automaticien spécifie et programme le système de commande, et celle pour laquelle le programme est – au moins partiellement – le produit d'un algorithme d'apprentissage automatique. Cette dernière voie peut répondre au nom de « automatisation apprise (*learned automation : LA*) ».

Dans le cas du LA, pour obtenir un programme de commande de la qualité requise par le cahier des charges, plusieurs critères doivent être pris en compte :

- l'adéquation de l'algorithme d'apprentissage avec les fonctions à réaliser;
- la quantité d'exemples à fournir à l'algorithme;
- la qualité de ces exemples (il peut y avoir des erreurs de labellisation ou une mauvaise représentativité des échantillons, etc.);
- et l'aptitude de l'algorithme d'apprentissage à fournir des indicateurs de ses propres performances.

Si l'évaluation de ces critères permet de répondre aux exigences des fonctions réalisées par la machine, le programme de commande aura un comportement tout à fait comparable à une machine conventionnelle. Son mode de programmation aura simplement fait appel à une IA; un moteur d'inférence (ou partie du système qui réalise la prise de décision) déterministe exécutera son programme.

Cependant, trois aspects de cet apprentissage méritent une attention particulière : la banque de données, l'opacité du programme LA et l'apprentissage continu.



## La banque de données

Comme cela a déjà été évoqué, les algorithmes d'apprentissage requièrent des exemples pour construire leur logique. Dans le cas de l'apprentissage supervisé, ces exemples sont constitués d'une information d'entrée – il s'agit de la question posée au moteur d'inférence – et du résultat attendu, soit la réponse. Cette désignation du résultat attendu est appelée « labellisation » (ou encore « étiquetage »). L'ensemble des données d'entrées associées à leur label constitue le jeu de données labellisées.

Pour apprendre à résoudre un problème complexe, le système d'IA devra analyser de nombreux couples de données [question; réponse]. Le nombre de couples de données pourra varier de plusieurs centaines à plusieurs centaines de milliers. Ainsi, la première difficulté est de constituer le jeu de données et de les labelliser. Généralement, la labellisation est effectuée par des opérateurs humains, ce qui constitue une tâche laborieuse. Elle peut être réalisée aisément pour des problèmes simples, mais elle peut également devenir titanesque quand des millions de données sont concernées. En outre, par construction, plus la tâche est complexe, plus le nombre

d'exemples nécessaires est élevé et plus la labellisation d'une donnée est compliquée, car celle-ci demande davantage d'acuité aux opérateurs.

À titre d'exemple, l'utilisation d'une machine capable de pulvériser de l'herbicide sur les adventices tout en épargnant les cultures<sup>2</sup> illustre cette difficulté. La machine utilise une caméra dotée d'un algorithme d'apprentissage pour discriminer les types de plantes et pilote les buses de pulvérisation de l'herbicide. L'entraînement de cette IA nécessite plusieurs dizaines de milliers d'images de plantes, et leur labellisation de réelles compétences en agronomie.

En plus des contraintes de labellisation, la grande masse de données nécessaire à l'entraînement des systèmes d'IA présente une autre difficulté : celle de la méthode pour constituer une telle banque de données.

La Figure 1 présente un logigramme des étapes nécessaires à la constitution d'une banque de données, ainsi que les contraintes (en orange) que cela implique.

Dans le cas où les données sont préexistantes, il est nécessaire de s'assurer de leur qualité et de leur représentativité du problème posé. Pour évaluer, voire améliorer ces données, un traitement

↑ FIGURE 1  
Logigramme de constitution d'une banque de données.



**ENCADRÉ 2  
EXEMPLE DE LABELLISATION  
APPLIQUÉ AU TRI DE MYRTILLES**

Pour créer une installation de contrôle qualité automatique de myrtilles, plusieurs centaines de milliers d'exemples peuvent être nécessaires afin de distinguer toute la diversité des critères qualifiant l'état des fruits (présence ou absence de défauts). Sur une installation de tri manuel, une machine d'acquisition d'images sur la bande transporteuse des myrtilles qualifiées de « bonnes », et une autre sur celles qualifiées de « mauvaises », permettent de constituer automatiquement la banque d'images labellisées. Il s'agit ici d'une labellisation *a priori*, car la classe de la myrtille est définie par l'emplacement de la prise de vue (bande « myrtilles sans défaut » ou bande « myrtilles avec défaut »).

Voir par exemple :  
<https://www.tomra.com/en/food/machines/kato-lucai>

Télé-opération d'un robot agile dans le cadre de l'étude INRS sur l'impact de l'IA dans les machines.

préalable peut s'avérer nécessaire. Ce traitement fait appel à des compétences particulières (*Data science*) encore peu répandues.

Si les données d'apprentissage ne sont pas disponibles, il faudra les « produire ». Pour ce faire, deux stratégies peuvent être mises en place :

- la production manuelle (cette organisation est incompatible avec des banques nécessitant d'énormes quantités de données) ;
- l'acquisition automatique de données, avec deux possibilités :
  - hors ligne : utilisation d'une machine spécifiquement dévolue à la construction de la banque de données ;
  - en ligne : une machine acquiert les données et les enregistre. Si cette acquisition est réalisable sans contrainte rédhibitoire pendant la production, elle n'engendrera pas d'indisponibilité de la machine.

Ce dernier mode – automatique, en ligne, en production – sera privilégié car il n'implique pas la réalisation de machine spécifique pour la construction de la banque de données, et pas d'arrêt de production. Ensuite, si les principes de production de données



© INRS/2023

ne génèrent pas intrinsèquement de labellisation, il sera nécessaire de les labelliser. Parce qu'elle est laborieuse lorsqu'elle est réalisée *a posteriori*, cette tâche sera préférentiellement effectuée pendant l'acquisition des données (lorsque c'est possible).

De plus, les erreurs de labellisation peuvent avoir des conséquences importantes sur la qualité des résultats du LA. Parce qu'elle ne nécessite pas de traitement manuel sur les données, une procédure d'étiquetage *a priori* générera moins d'erreurs.

## L'opacité du programme LA

Dès lors que la banque de données est constituée et que l'algorithme d'apprentissage est sélectionné, il convient de lancer l'entraînement du modèle qui consiste donc à faire « digérer » les données à l'algorithme. Cette procédure informatique, chronophage et mobilisatrice en ressources système, produit le modèle entraîné qui constituera le moteur du LA. Elle fournira également un score de performance du système qu'il est nécessaire de construire.

Toujours dans le cas d'un apprentissage supervisé pour résoudre un problème de classification (Cf. Encadré 2 pour la classification de fruits), les indicateurs de performance sont construits à partir d'une matrice, appelée matrice de confusion (Cf. Encadré 3).

Plusieurs indicateurs (ou métriques) peuvent être déduits de la matrice de confusion, comme la justesse (les 89 % de l'exemple cité en Encadré 3), la précision, la sensibilité, la spécificité, etc. Tous ces indicateurs permettent d'améliorer la lecture de la matrice de confusion, mais aucun n'apporte d'information sur la construction de chaque décision de l'IA (pourquoi 6 défauts de myrtilles n'ont-ils pas été observés ? Pourquoi 5 myrtilles ont-elles été considérées avec défaut, alors qu'elles n'en présentaient pas ?...).

Ce problème « d'opacité de la prise de décision des systèmes LA » est la conséquence de la multitude des paramètres (plusieurs centaines de milliers) réglés par l'algorithme d'apprentissage lors de l'entraînement. Même si ces valeurs sont toutes accessibles, la complexité engendrée par leur multiplicité rend impossible la relecture du cheminement d'une décision.

Ainsi, parce que les tests ne peuvent être exhaustifs (impossible de présenter au système toutes les myrtilles dans toutes les conditions d'observation) et parce qu'il est impossible de suivre le chemin qui a conduit à la décision prise par le système, les systèmes LA restent opaques et leur validation<sup>3</sup> s'avère à ce jour inatteignable.

## L'apprentissage continu

Dans le cadre de l'apprentissage supervisé, les systèmes LA sont programmés grâce à des

### ENCADRÉ 3

#### LA MATRICE DE CONFUSION

Cette matrice est renseignée à partir d'un jeu de données labellisées qui n'ont pas été utilisées pour l'entraînement. Ces données sont présentées au système pour tests. Les résultats de ces tests permettent de renseigner, par dénombrement, cette matrice. Grâce au jeu de tests, elle est générée automatiquement par l'algorithme d'apprentissage.

		RÉSULTAT DE LA CLASSIFICATION (PRÉDICTION)	
		Positif	Négatif
RÉALITÉ (LABEL)	Positif	<b>VRAI</b> Positif	<b>FAUX</b> Négatif
	Négatif	<b>FAUX</b> Positif	<b>VRAI</b> Négatif

Pour illustrer, prenons un jeu de tests comprenant 100 myrtilles dont 21 comportent un défaut.

Le système renvoie la matrice de confusion suivante :

		RÉSULTAT	
		Sans défaut	Avec défaut
MYRILLES	Sans défaut	<b>VP</b> <b>74</b>	<b>FN</b> <b>5</b>
	Avec défaut	<b>FP</b> <b>6</b>	<b>VN</b> <b>15</b>

Cette matrice montre donc que sur 100 myrtilles :

- 89 ont été bien évaluées (« vrais positifs et vrais négatifs ») ;
- 5 ont été perçues comme présentant des défauts alors qu'elles n'en ont pas en réalité (« faux négatifs ») ; et
- 6 ont des défauts avérés mais qui sont passés inaperçus (« faux positifs »).

Elle présente un taux de réussite au test global de 89 %.

Elle permet également de distinguer les types d'erreur :

- 5 myrtilles sans défaut iront rejoindre les déclassées (ce qui génère une perte de productivité) ;
- 6 myrtilles présentant des défauts seront vendues comme de qualité supérieure (ce qui génère une perte de qualité).



exemples de leur propre tâche. Ainsi, après leur mise en service et donc leur entraînement initial, ils seront en mesure d'observer les nombreux exemples qu'eux-mêmes produiront. Il paraîtrait donc naturel d'utiliser cette source de nouveaux exemples pour enrichir leur programme et ainsi les faire progresser. On pourrait appeler ce principe « l'apprentissage continu ».

Or, les résultats enregistrés par un système en service proviennent de lui-même, ils ne peuvent donc pas être considérés comme fiables. En effet, le jeu d'exemples utilisé pour l'entraînement doit être labellisé (le résultat attendu doit être attaché à l'observation). En d'autres termes, sans validation de chaque exemple par un tiers, le jeu de données autoproduit par le système ne peut le faire progresser.

Par contre, une procédure d'optimisation, supervisée par un opérateur humain et dans un mode spécifique est tout à fait envisageable. Pour reprendre l'exemple précédent, un opérateur corrigeant en sortie de machine les erreurs d'appréciation du système permettrait de labelliser les faux positifs et les faux négatifs pour constituer un nouveau jeu de données d'apprentissage.

### Conclusion

Grâce à l'IA, une machine acquiert de nouvelles capacités, comme le classement d'objets naturels ou artificiels (cultures ou adventices, myrtilles, déchets, etc.) selon certaines caractéristiques. Ces facultés étaient inaccessibles aux machines dépourvues d'IA. Cependant, l'IA n'affecte pas directement le fonctionnement de la machine; par exemple, celle-ci ne prendra pas l'initiative de démarrer spontanément ou de partir sur des chemins qui lui sont inconnus. En somme, si les machines dotées d'IA ont des capacités supérieures à celles qui n'en ont pas, elles ne sont pas plus autonomes pour autant. L'IA, dans une machine, n'implique pas un comportement « intelligent » de celle-ci.

Les machines dont le système intègre de l'automatisme appris (LA) restent des machines, avec leurs énergies et leurs dangers. Cette particularité du système n'engendre pas de danger supplémentaire. Elle permet de créer des automatismes plus complexes, capables de traiter des informations plus diversifiées, comme des images d'objets naturels ou de milieux extérieurs ouverts. Cependant, la nature opaque de l'apprentissage automatique rend la validation des programmes extrêmement complexe, voire irréalisable. C'est la raison pour laquelle le LA ne peut jusqu'alors pas intégrer la chaîne logique d'une fonction dédiée à la sécurité d'une machine.

Pourtant, la réglementation européenne, dans le règlement « Machines », qui remplacera l'actuelle

directive 2006/42/CE en janvier 2027<sup>4</sup>, prévoit cette éventualité. Ce règlement prendra en compte les machines dont une fonction de sécurité utilisera du LA. De plus, une autre mention à l'apprentissage automatique est faite : il envisage des machines dont le comportement serait auto-évolutif, c'est l'apprentissage continu. Il s'agirait de machines dont l'entraînement continuerait pendant leur exploitation.

La démarche d'évaluation et de réduction des risques reste applicable pour les machines intégrant l'automatisme appris dans leurs fonctions standards. En revanche, pour les fonctions de sécurité, même si la future réglementation l'envisage, l'inexplicabilité des algorithmes reste aujourd'hui un obstacle à la certification « CE » des machines. ●

1. Voir : <https://dictionnaire.lerobert.com/definition/autonomie>

2. Voir : <https://carbonbee-agtech.fr/>

3. Selon la norme NF EN 61508-4, la validation est la confirmation, par examen et apport de preuves tangibles, que les exigences particulières pour un usage spécifique prévu sont satisfaites.

4. Voir dans ce numéro : dossier « Machines », pp. 18-44. Accessible sur : <https://www.inrs.fr/publications/hst/dossier.html>

### POUR EN SAVOIR +

• ISO – Document ISO/IEC TR 29119-11 – Ingénierie du logiciel et des systèmes – Essais du logiciel – Partie 11 : lignes directrices relatives aux essais portant sur les systèmes dotés d'IA, 2020.

Accessible sur : <https://www.iso.org/fr/standard/79016.html> (document payant).

• INRS – L'intelligence artificielle au service de la santé et de la sécurité au travail : enjeux et perspectives à l'horizon 2035. Synthèse. Coll. Veille & prospective, 2022, PV 20. Accessible sur : <https://www.inrs.fr/media.html?refINRS=PV%2020>

• INRS/EUROGIP – L'intelligence artificielle rencontre la santé et sécurité au travail. In : Journée Euroshnet, Paris, 20 octobre 2022. Accessible sur : <https://eurogip.fr/conference-euroshnet-2022-intelligence-artificielle-et-sst/> ; et <https://www.inrs.fr/media.html?refINRS=NO%2037>

• ISO – Document TR 22100-5 – Sécurité des machines – Relation avec l'ISO 12100 – Partie 5 : implications de l'intelligence artificielle pour l'apprentissage automatique, 2021. Accessible sur : <https://www.iso.org/fr/standard/80778.html> (document payant).