eCH-0272 – Transparence, explicabilité et risques des systèmes d'IA

Nom	Transparence, explicabilité et risques des systèmes d'IA
eCH-nombre	eCH-0272
Catégorie	Norme
Stade	Défini
Version	1.0.0
Statut	Approuvé
Date de décision	2025-06-18
Date de publication	2025-04-08
Remplace la version	Nouveau – Sélectionnez un élément.
Conditions préalables	---
Annexes	---
Langues	Allemand (original), français (traduction)
Groupe spécialisé	IA
Éditeur / distribution	Association eCH, Räffelstrasse 20, 8045 Zurich T 044 388 74 64 / info@ech.ch / www.ech.ch

Condensé

La présente norme constitue une norme conceptuelle qui définit les exigences en matière de transparence, d'explicabilité et de risques relatifs aux systèmes d'IA. Partant d'une analyse approfondie des conditions-cadres, des méthodes élaborées et des résultats obtenus, cette norme sert de base aux autres normes, fruits des travaux du groupe spécialisé IA d'eCH.

Table des matières

1 Introduction 4
1.1 Statut 4
1.2 Champ d'application 4
1.3 Champ d'application de la norme 6
1.4 Procédure 7
1.5 Notation de la norme 8
1.6 Cycle de vie de la norme 9
2 Principes juridiques fondamentaux 9
2.1 Lois et ordonnances en Suisse 9
2.2 Lois et ordonnances dans l'UE/l'EEE 11
2.2.1 Règlement de l'UE sur l'IA (AI Act) réglementant l'intelligence artificielle 11
2.2.2 Convention-cadre du Conseil de l'Europe sur l'IA 12
3 Glossaire 13
4 Principes concernant l'IA en Suisse 16
4.1 «Directives d'IA pour l'administration fédérale» 16
4.2 Code de bonnes pratiques de la Confédération en matière de science des données et d'espaces de données dignes de confiance 17
5 Phases du cycle de vie et rôles des Stakeholders des systèmes d'IA 17
6 Vue d'ensemble des normes et standards internationaux 19
7 Objectifs de protection et d'acceptation 20
8 Exigences relatives aux systèmes d'IA 24
8.1 Généralités 26
8.2 Droits fondamentaux 32
8.3 Protection des données 34
8.4 Droits d'auteur 37
8.5 Autonomie 40
8.6 Équité 45
8.7 Obligation de rendre des comptes 48

8.8 Transparence 51
8.9 Explicabilité 55
9 Aide à la mise en œuvre & exemples d'application 64
10 Checklist pour la mise en œuvre d'eCH-0272 66
11 Adéquation à la pratique des normes en matière d'IA 67
12 Sécurité 67
13 Exclusion de responsabilité – droits de tiers 67
14 Droits d'auteur 68
Annexe A – Références & bibliographie 69
Annexe B – Collaboration & vérification 70
Annexe C – Abréviations et glossaire 70
Annexe D – Modifications par rapport à la version précédente 71
Annexe E – Liste des illustrations 71
Annexe F – Liste des tableaux 71

Remarque

La formulation employée dans le présent document pour désigner les personnes est neutre en termes de genre. Elle repose sur le guide de la Chancellerie fédérale. On recourt, selon la situation, à des doublets intégraux (citoyens et citoyennes), à des formes abstraites en termes de genre (personne assurée), à des formes neutres du point de vue du genre (les assurés) ou à des périphrases dépourvues de référence à la personne. L'utilisation du masculin générique (citoyens) n'est plus admise. Les formes intégrales sont employées dans les textes continus, autrement dit les textes constitués de phrases rédigées. Les formes abrégées sont acceptées dans les passages de texte concis, les tableaux par exemple. On utilise alors la forme courte avec barre oblique, toutefois sans tiret (référent/e). Les points médians et autres caractères similaires sont proscrits.

1 Introduction

1.1 Statut

Approuvé: le document a été approuvé par le Comité des experts. Il a pouvoir normatif pour le domaine d'utilisation défini dans le domaine de validité donné.

1.2 Champ d'application

La présente norme spécifie les exigences minimales en matière de transparence, d'explicabilité et de risques relatifs aux systèmes d'IA. Cette norme est en ceci intéressante qu'elle propose une définition structurée des exigences fondées sur des objectifs de protection et d'acceptation clairement énoncés. Elles garantissent le respect des lois en vigueur, telles que la loi sur la protection des données ou le droit d'auteur (voir chapitre 2), préservent les principes éthiques et favorisent la confiance dans l'utilisation des systèmes d'IA. Un juste équilibre entre innovation et conformité contribue à instaurer des espaces de liberté technologiques sans mettre en péril les droits fondamentaux, la transparence ou la traçabilité. La norme soutient les institutions publiques et privées dans leurs efforts de mise en œuvre et de surveillance des systèmes d'IA, contribuant de fait à renforcer la capacité d'innovation et la compétitivité de la Suisse.

Figure 1: Équilibre entre innovation et conformité

En outre, la norme n'émet délibérément aucune préconisation quant aux technologies, implémentations ou architectures système spécifiques. Elle a bien plus vocation à s'appliquer à toutes les organisations et à fournir des repères pour une utilisation responsable des systèmes d'IA, quelles que soient les solutions logicielles ou les fournisseurs.

La norme s'appuie sur celles publiées par l'association eCH en Suisse afin de définir des règles présidant au développement, à l'utilisation et à la documentation de systèmes d'IA, en tenant compte des références normatives propres à l'administration fédérale suisse (en particulier les lignes directrices «Intelligence artificielle» pour la Confédération), mais également d'autres lois applicables au cas par cas (par ex. le règlement européen sur l'IA, l'AI-Act de l'UE). Outre une analyse rigoureuse du cadre juridique en place, les normes internationales ont elles-aussi été examinées aux fins d'élaborer la présente norme (voir chapitre 6). La norme eCH-0272 se distingue des normes internationales sur les points suivants:

- Spécificités suisses: la présente norme tient compte des lois et conditions-cadres propres à la Suisse, dont les normes internationales ne traitent pas en détail.
- Application pratique: l'applicabilité pratique au contexte suisse y tient une place centrale, avec notamment des préconisations spécifiques à l'attention des institutions publiques et privées.
- Normes ISO: les normes ISO (ISO/IEC 22989 et ISO/IEC 23053) sont axées sur les exigences de transparence et d'explicabilité, non sans fournir une orientation générale pour un grand nombre d'applications et de secteurs. Ces normes s'inscrivent dans une approche générale partant des risques afin d'identifier et d'évaluer les risques liés à la mise en œuvre de systèmes d'IA.
- eCH-0272: met en avant la transparence et l'explicabilité des systèmes d'IA, avec des exigences et des préconisations spécifiques adaptées aux besoins et aux règles légales de la Suisse. La norme eCH-0272 suit également une logique fondée sur les risques, tout en définissant des mesures propres à la législation suisse.

La norme eCH-0272 doit être utilisée en tant que norme de base conceptuelle, qui pose un socle solide pour les autres normes eCH en matière d'IA. Elle offre une vue d'ensemble robuste de l'environnement de normalisation de l'IA et illustre la façon d'assurer la traçabilité (Traceability) des résultats.

Contrairement à un White Paper, qui propose en général une analyse discursive ou des recommandations d'action, eCH-0272 définit des exigences concrètes, fondées sur les risques, en matière de transparence, d'explicabilité et de gestion des risques inhérents aux systèmes d'IA.

La norme permet une mise en œuvre structurée grâce à:

- des exigences formulées de manière claire, tout au long du cycle de vie d'un système d'IA
- le rapprochement avec des objectifs concrets de protection et d'autres ayant trait à l'acceptation
- l'orientation par rapport à des normes internationales (par ex. ISO/IEC 22989, 42001) en mettant l'accent sur les conditions-cadres suisses.

eCH-0272 est ainsi synonyme de valeur ajoutée opérationnelle pour les organisations, autorités et prestataires de services qui cherchent à évaluer, à acquérir ou à exploiter des systèmes d'IA de manière responsable.

1.3 Champ d'application de la norme

La présente norme s'adresse en priorité à la Confédération, aux cantons et aux communes. Les établissements de droit public ainsi que les entreprises du secteur privé peuvent eux aussi l'appliquer.

Les groupes cibles précis sont décrits comme suit:

Fournisseur: toute personne physique ou morale, autorité publique, institution ou autre entité qui conçoit ou fait concevoir un système d'IA en vue de le commercialiser ou de le mettre en service en son nom propre ou sous sa marque, à titre gracieux ou onéreux.

Utilisateur: toute personne physique ou morale, autorité publique, institution ou autre entité qui utilise un système d'IA sous sa responsabilité propre. Il faut pour ce faire que le système d'IA soit utilisé exclusivement dans le cadre d'une activité personnelle et non professionnelle.

Autorités de régulation: autorités qui supervisent les systèmes d'IA ou qui contrôlent la conformité avec la législation des systèmes d'IA et sont habilitées à émettre des règlements et des prescriptions.

D'autres rôles de Stakeholders figurent également au chapitre 5. Ceux-ci se rapportent à la phase du cycle de vie des systèmes d'IA.

Délimitation: la norme eCH-0272 porte sur les systèmes d'IA pré-entraînés et exploités localement. Les systèmes GenAI basés sur le cloud et bénéficiant d'un training en continu, qui fonctionnent par ex. sur des plateformes telles qu'OpenAI, AWS ou Google Cloud, échappent au champ d'application de la présente norme, en raison des exigences supplémentaires impliquées en matière de protection des données et de sécurité.

1.4 Procédure

La procédure de développement de la norme eCH-0272 suit un processus séquentiel censé garantir que toutes les exigences pertinentes sont prises en compte de manière structurée et globale. Le processus comporte les étapes suivantes:

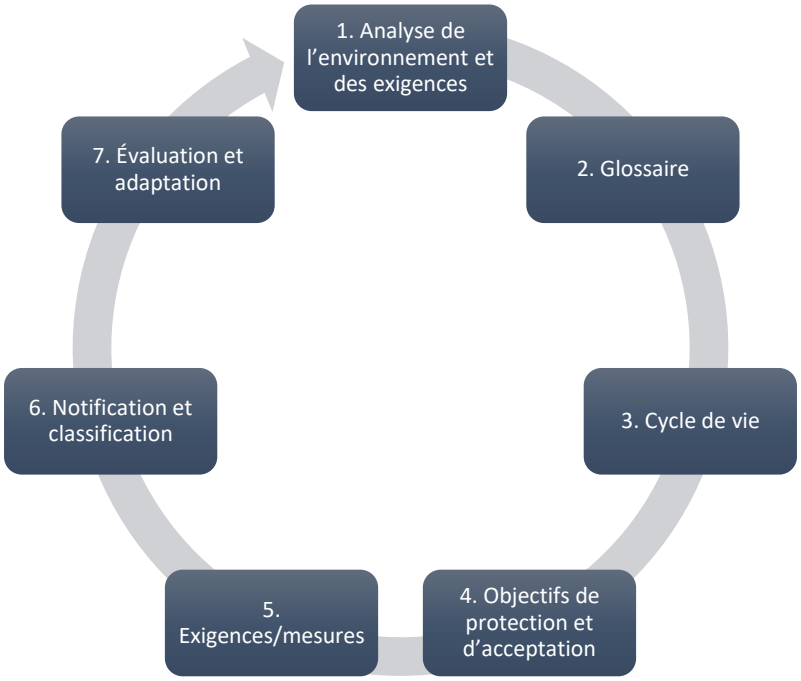

Figure 2: Processus de la norme eCH-0272

1. Analyse de l'environnement et des exigences: Il est en premier lieu procédé à une analyse exhaustive des besoins, visant à identifier toutes les conditions-cadres et corpus de règles applicables aux systèmes d'IA. Une autre analyse, de l'environnement cette fois, est effectuée en parallèle en tenant compte des normes nationales et internationales, des progrès technologiques actuels et des attentes de la part de la société. Cette analyse sert de base à la définition des exigences de transparence, d'équité et d'autonomie du système d'IA et permet de s'assurer que le système d'IA répond bien tant aux exigences réglementaires qu'aux attentes éthiques et technologiques en vue de favoriser l'innovation.

2. Glossaire: Le glossaire est défini comme une rubrique à part entière afin de garantir que l'intégralité des termes utilisés le sont avec cohérence et précision tout au long du processus. Un glossaire précis garantit que toutes les étapes du processus sont bien comprises et mises en œuvre de façon homogène et que les exigences et les mesures de protection sont interprétées de manière cohérente par toutes les personnes impliquées.

3. Détermination du cycle de vie: L'étape suivante consiste à définir les phases du cycle de vie du système d'IA, qui sont définies dans la norme en tant qu'initialisation, conception & développement, vérification & validation, implémentation, exploitation & surveillance, réévaluation et mise hors service, qui s'appuient sur une norme ISO. Les exigences et les mesures de protection qui s'imposent sont définies pour chacune des phases. Qui plus est, tous les rôles pertinents des Stakeholders, tels que fournisseurs, utilisateurs et autorités de régulation, sont identifiés. L'opération consiste à déterminer le rôle de chaque Stakeholder dans le cycle de vie du système d'IA et les informations dont ils ont besoin.

4. Définition des objectifs de protection et d'acceptation: Les objectifs de protection et d'acceptation sont définis à partir des analyses précédentes. Ces objectifs sont axés sur les droits fondamentaux, la protection des données, l'autonomie, etc. (objectifs de protection) ainsi que sur la confiance dans les systèmes d'IA et leur acceptation au sein de la société (objectifs d'acceptation). Les objectifs de protection visent les risques potentiels tels que la discrimination, tandis que les objectifs d'acceptation cherchent à promouvoir la transparence et l'explicabilité, renforçant ainsi la confiance des utilisatrices et utilisateurs et des Stakeholders.

5. Exigences et mesures: Des exigences et des mesures de protection détaillées sont élaborées sur la base de ces objectifs de protection et d'acceptation. Les analyses des biais, les mécanismes de transparence et la documentation obligatoire en sont de bons exemples. Les mesures comprennent des dispositions d'ordre technique et organisationnel censées garantir le respect permanent des exigences fixées.

6. Évaluation après notification et classification: Une évaluation reposant sur les risques est effectuée pour chacune des exigences. Le risque est ainsi classé selon différentes catégories (faible, moyen, élevé) et les objectifs de protection et d'acceptation sont définis en conséquence. Les risques tels que la discrimination ou les problèmes d'équité font l'objet d'une vérification ciblée.

7. Évaluation et adaptation: Les exigences et mesures sont contrôlées en permanence et adaptées, le cas échéant, en fonction des expériences acquises et des retours d'information reçus.

1.5 Notation de la norme

Les exigences mentionnées dans les normes sont spécifiées selon la terminologie [RFC2119], les expressions suivantes, qui sont mises en évidence par des mots écrits en MAJUSCULE avec les significations suivantes, y sont utilisées:

IMPÉRATIF: la personne responsable doit réaliser l'objectif.

RECOMMANDÉ: la personne responsable peut renoncer à la réalisation de l'objectif pour des motifs importants.

FACULTATIF: la personne responsable est libre de choisir si elle souhaite réaliser l'objectif.

1.6 Cycle de vie de la norme

Cette norme peut être modifiée à tout moment par les utilisateurs par le biais d'une demande en ce sens selon la norme eCH003. En outre, cette norme doit être réexaminée dans sa globalité une fois tous les deux ans afin de s'assurer qu'elle répond bien aux besoins actuels des utilisateurs.

2 Principes juridiques fondamentaux

À l'heure où cette norme est élaborée, il n'existe en Suisse aucune législation régissant spécifiquement l'intelligence artificielle (IA). On trouve cependant des lois et des règlements (la loi fédérale sur la protection des données par exemple) dont les dispositions couvrent également l'IA.

Au sein de l'Union européenne (ci-après l'UE), le règlement sur l'intelligence artificielle (AI Act) [1] est entrée en vigueur le 1er août 2024. Les premières interdictions pour les systèmes d'IA dans certains cas de figure sont entrées en vigueur le 2 février 2025.

La liste ci-dessous des principes légaux en Suisse et dans l'UE n'est pas exhaustive.

2.1 Lois et ordonnances en Suisse

Constitution fédérale de la Confédération suisse (Cst., RS 101)

Les droits fondamentaux (dont la dignité humaine, la protection de la sphère privée et la protection de la personnalité) sont stipulés aux articles 7 et suivants de la Constitution fédérale et doivent être respectés dans l'ordre juridique général. Il faut, au moment de la conception et de l'utilisation de l'IA, prendre en compte les considérations relatives aux droits fondamentaux et à l'éthique («ethics by design»). Le recours à l'intelligence artificielle ne doit pas bafouer la dignité humaine. Le principe de l'égalité de traitement énoncé à l'article 8 de la Constitution fédérale (Cst., RS 101) revêt une importance majeure, notamment pour les systèmes d'IA. Ces systèmes présentent le risque d'aggraver les inégalités sociales existantes ou de conduire à de nouvelles formes de discrimination, en raison par exemple de données de training biaisées, de modèles non représentatifs ou d'une logique de décision opaque.

Loi fédérale sur la protection des données (loi sur la protection des données, LPD; RS 235.1)

Cette loi vise à protéger la personnalité et les droits fondamentaux des personnes physiques dont les données à caractère personnel sont traitées. La LPD est directement applicable à la mise en œuvre de traitements de données basés sur l'IA2. Les concepteurs, fournisseurs et utilisateurs de systèmes d'IA sont tenus de se conformer à la LPD dès lors qu'ils traitent des données à caractère personnel, comme le devoir d'information en cas de décision individuelle automatisée ou l'exécution d'une analyse d'impact relative à la protection des données lors d'un traitement de données présentant un

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

Communiqué du Préposé fédéral à la protection des données du 9 novembre 2023.

risque élevé pour les droits de la personnalité ou les droits fondamentaux des personnes concernées par exemple. Ce risque élevé est inhérent à la nature, à l'étendue, aux circonstances et à la finalité du traitement, notamment en cas d'utilisation de nouvelles technologies. **Loi fédérale sur le droit d'auteur et les droits voisins (loi sur le droit d'auteur, LDA; RS 231.1)**

La loi sur le droit d'auteur protège la propriété intellectuelle d'un auteur. Concernant l'IA, la question de la propriété intellectuelle importe également dans le cadre du traitement ou de l'utilisation d'œuvres couvertes par le droit d'auteur. Concernant l'utilisation de contenus par des systèmes d'IA, la Suisse n'a encore défini aucune évaluation juridiquement contraignante des droits d'auteur. On peut certes imaginer que le training de systèmes d'IA et l'élaboration d'outputs d'IA donnent lieu à des utilisations d'œuvres soumises au droit d'auteur. Toutefois, il n'existe pour l'heure aucune évaluation contraignante à ce sujet.

Code pénal suisse (CP; RS 311.0) et autres lois applicables

Les systèmes d'IA n'ayant pas de personnalité juridique, tout préjudice causé par l'IA engage la responsabilité d'une personne physique ou morale, dès lors que les conditions de responsabilité sont remplies. En pareil cas, le CP ainsi que le Code des obligations (CO ; RS 220) ou la Loi fédérale sur la responsabilité du fait des produits (Loi sur la responsabilité du fait des produits, LRFP ; RS 221 112 944) entre autres s'appliquent.

LMETA – Loi fédérale du 17 mars 2023 sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités (RS 172.075):

La LMETA impose aux autorités fédérales d'effectuer leurs tâches de façon transparente, compréhensible et efficace en recourant à des moyens électroniques. L'utilisation de systèmes d'IA dans l'administration implique dès lors que les processus décisionnels soient conçus de manière compréhensible et vérifiable et que les utilisateurs disposent d'outils d'aide adéquats.

Les points suivants sont particulièrement pertinents à cet égard:

- Art. 1 LMETA: But (prestations des autorités électroniques efficaces, sûres et conviviales)
- Art. 4 LMETA: Règles visant la confiance numérique
- Art. 9 LMETA: Exigences relatives aux procédures électroniques, y compris l'automatisation et l'interopérabilité

2.2 Lois et ordonnances dans l'UE/l'EEE

Dans le cadre du développement de cette norme, les lois et ordonnances pertinentes au sein de l'UE ont été analysées et examinées en détail. Ce faisant, nous avons fait en sorte que divers aspects de ces textes soient repris dans les différentes thématiques traitées par la norme. Il convient toutefois de noter que seules les dispositions de ces lois et ordonnances, dont la conformité avec les normes suisses est établie et qui ont été jugées propres à répondre aux exigences et besoins spécifiques de la présente norme, y ont été intégrées. La compatibilité et la pertinence des réglementations par rapport au contexte suisse président donc à leur intégration. Un traitement exhaustif et approfondi de ces réglementations internationales être prévu dans les versions à venir de la norme.

2.2.1 Règlement de l'UE sur l'IA (AI Act) réglementant l'intelligence artificielle

Le règlement sur l'IA de l'UE pose un cadre juridique global qui fixe des règles et des normes claires concernant le recours à l'intelligence artificielle, dans un souci de protection des droits fondamentaux. Ces règles garantissent que les systèmes d'IA respectent la dignité, la vie privée et les autres droits fondamentaux des personnes et ne sont pas discriminatoires, manipulateurs ou préjudiciables, de quelque manière que soit.

Pour améliorer la confiance dans les technologies d'IA, le règlement sur l'IA de l'UE exige la traçabilité et l'explicabilité des systèmes d'IA.

Le règlement opère une classification en quatre catégories de risque: faible, limité, élevé et inacceptable. Les systèmes d'IA présentant un risque inacceptable sont proscrits. Les exigences légales varient selon la classe de risque, les systèmes présentant un risque élevé étant soumis à des règles particulièrement contraignantes. Les fournisseurs de tels systèmes d'IA ont l'obligation de procéder à une évaluation exhaustive des risques.

Il impose en outre des obligations de documentation et le respect de normes techniques spécifiques. Ce règlement a pour finalité un meilleur fonctionnement du marché intérieur en posant un cadre juridique homogène régissant notamment le développement, la commercialisation et l'utilisation de l'intelligence artificielle, en accord avec les valeurs défendues par l'UE.

Tous les acteurs tombant sous le coup du règlement de l'IA devraient développer et déployer des systèmes d'IA à risque élevé en accord avec les six «principes d'IA» suivants (Commission européenne, 2023):

- Action humaine et contrôle humain: Les systèmes d'IA doivent être au service de l'humain, respecter sa dignité et son autonomie personnelle, et fonctionner de manière à pouvoir être contrôlés et surveillés par l'humain.
- Robustesse technique et sécurité: Les dommages involontaires et inattendus devraient être réduits au minimum et les systèmes d'IA faire preuve d'une certaine robustesse en cas de problèmes involontaires.
- Data Governance: Les systèmes d'IA à risque élevé s'appuyant sur des techniques de training de modèles à partir de données, les jeux de données de training, de validation et de test doivent être pertinents, représentatifs, sans erreurs et complets.
- Transparence: La traçabilité et l'explicabilité doivent être possibles et les personnes avoir bien conscience qu'elles sont en train d'interagir avec un système d'IA.

- Diversité, non-discrimination et équité: Les systèmes d'IA doivent impliquer différents acteurs et promouvoir l'égalité d'accès, l'égalité des sexes et la diversité culturelle, et inversement, éviter les répercussions discriminatoires.
- Bien-être sociétal et environnemental: Les systèmes d'IA doivent être durables et respectueux de l'environnement, et être développés et utilisés pour le bien de tous les êtres humains.

2.2.2 Convention-cadre du Conseil de l'Europe sur l'IA

La Convention-cadre du Conseil de l'Europe sur l'IA vise à négocier le premier accord international contraignant au monde portant sur l'IA. Cette convention s'appuie sur les valeurs fondamentales défendues par le Conseil de l'Europe – droits de l'homme, démocratie et État de droit – et entend s'assurer de leur respect dans le contexte de l'IA également. Outre la protection de ces valeurs, la convention encourage également l'innovation dans le développement et l'application des technologies d'IA.

Cette convention a été signée par Andorre, la Géorgie, l'Islande, la Norvège, la République de Moldavie, Saint-Marin, le Royaume-Uni, Israël, les États-Unis et l'Union européenne. Une ratification par la Suisse impose de transcrire l'accord dans le droit national.

La convention revêt pour la Suisse une importance particulière, car elle vient poser un cadre juridiquement contraignant qui garantit que les systèmes d'IA développés et déployés en Suisse répondent aux normes internationales les plus strictes. Ces normes portent notamment sur la transparence, la robustesse, la non-discrimination et la protection de la vie privée, autant d'aspects particulièrement importants pour la Suisse, pays où l'innovation et la technologie occupent une place prépondérante. Qui plus est, par sa participation à cette convention, la Suisse assoit davantage sa position de pionnière dans le domaine de la réglementation d'une IA éthique et sûre au niveau international. Cela permet non seulement de promouvoir la confiance dans les technologies d'IA, mais également de soutenir la protection des droits fondamentaux dans un environnement marqué par une numérisation croissante.

La Convention-cadre commence par énoncer les principes applicables aux activités effectuées dans le cadre du cycle de vie des systèmes d'IA. Elle fixe également les possibilités de recours et le dispositif encadrant l'évaluation et l'atténuation des risques et des conséquences négatives. La mise en œuvre de la convention-cadre est également détaillée et assortie d'un mécanisme de suivi et de coopération ainsi que de dispositions finales.

3 Glossaire

Le glossaire dans le présent document a vocation à décrire les termes importants dans le contexte de l'intelligence artificielle (IA) qui sont pertinents pour la compréhension et l'application pratique de cette norme. Ces définitions proviennent d'un grand nombre de sources reconnues afin de garantir à la fois la clarté et la pertinence. Cette terminologie contribue à une interprétation cohérente des exigences et facilite l'application des objectifs de protection et d'acceptation définis dans la norme.

Les sources correspondantes sont précisées dans l'annexe A – Références & bibliographie (liens complets inclus) et permettent d'approfondir l'étude des différents concepts.

Terme	Définition/description
Intelligence artificielle (IA)	Systèmes qui exécutent des tâches faisant habituellement appel à l'intelligence humaine, telles que la perception visuelle, le traitement du langage ou la prise de décision.
Système d'IA	Un système d'IA est un système assisté par ordinateur qui, avec un certain degré d'autonomie, peut générer des prédictions, des recommandations ou des décisions ayant une influence sur des environnements physiques ou virtuels, pour des objectifs explicites ou implicites. Remarque: La norme eCH-0272 recourt à une large définition des «systèmes d'IA» selon l'OCDE. Celle-ci comprend les systèmes symboliques (par ex. les systèmes experts basés sur des règles), les systèmes statistiques déclaratifs (par ex. les modèles de régression classiques), les systèmes d'apprentissage automatique (par ex. les classificateurs supervisés, le Deep Learning) et les systèmes d'IA génératifs, pour autant qu'ils fassent l'objet d'un pré-training et d'une exploitation locale. Une telle diversité induit des différences de propriétés techniques, de contextes d'utilisation et de potentiels de risque. Une hétérogénéité dont tient compte la norme en optant pour une approche axée sur les risques (voir chapitre 8) et préconise une adaptation des exigences ciblée sur le type de système, le contexte d'utilisation et la dimension critique. Dans la pratique, une distinction entre les différents types de systèmes, tels que «système symbolique», «système statistique déclaratif», «système basé sur le ML», « IA générative (locale, pré-entraînée)» ou «système hybride» est par conséquent judicieuse.
Apprentissage automatique (ML)	Un domaine de l'IA qui s'intéresse aux algorithmes chargés d'analyser les données et capables d'améliorations adaptatives pour permettre des prédictions et des classifications.

Terme	Définition/description
GENAI	L'IA générative, parfois appelée Gen AI, est une intelligence artificielle (IA) qui génère un contenu original comme du texte, des images, de la vidéo, de l'audio ou du code logiciel en réponse au prompt ou à la demande d'un utilisateur. Remarque: Cette norme couvre uniquement les systèmes d'IA génératifs exploités localement et utilisés à partir de modèles pré-entraînés (par ex. LLM open source, générateurs d'images sans connexion cloud). Ces systèmes entrent dans le champ d'application de la présente norme dans la mesure où ils sont utilisés dans des environnements de système fermés et ne font l'objet d'aucun training ultérieur continu. Les systèmes GenAI basés sur le cloud, entraînés de manière dynamique et dotés d'une infrastructure contrôlée en externe (par ex. OpenAI GPT, Amazon Bedrock, Google PaLM API) ne sont pas concernés.
Système symbolique	Un système d'IA fondé sur la manipulation de symboles et de logiques explicites basées sur des règles. Les conclusions sont tirées de manière déterministe.
Système statistique déclaratif	Un système qui utilise des méthodes statistiques classiques afin d'identifier des modèles ou faire des prédictions – sans apprentissage adaptatif.
Système hybride	Un système d'IA qui combine des méthodes symboliques et d'apprentissage afin de mettre à profit les avantages des deux approches.
Transparence	Transparence: Propriété d'un système ou d'un processus dont les procédés et décisions sont compréhensibles et intelligibles pour des personnes extérieures.
Explicabilité	Mesure dans laquelle les mécanismes internes d'un système ou modèle peuvent être rendus compréhensibles pour des personnes.
Autonomie	Capacité d'un système à agir sans contrôle ni intervention extérieure.
Équité	Propriété d'un système dans lequel il n'y a ni désavantage ni préférence systématique vis-à-vis de groupes particuliers.
Obligation de rendre des comptes	Obligation faite à des personnes ou organisations d'assumer la responsabilité de leurs actes et des conséquences qui en résultent, et de les expliquer.
Objectifs de protection	Objectifs garantissant la protection par les systèmes d'IA des droits fondamentaux, de la vie privée et de l'autonomie des utilisatrices et utilisateurs, y compris la protection de la propriété intellectuelle.

Terme	Définition/description
Objectifs d'acceptation	Mesures visant à promouvoir la confiance et l'acceptation par la société des systèmes d'IA via la transparence et une perception positive de la part des utilisatrices et utilisateurs et des Stakeholders.
Modèle White Box	Modèle dans lequel les procédés internes et les processus de décision sont totalement visibles et compréhensibles pour l'utilisatrice ou l'utilisateur.
Modèle Black Box	Modèle dans lequel les procédés internes et les processus de décision ne sont pas visibles ou compréhensibles pour l'utilisatrice ou l'utilisateur.
SHAP (SHapley Additive exPlanations)	Fondé sur les valeurs de Shapley issues de la théorie des jeux, SHAP attribue à chaque caractéristique d'un modèle une importance en fonction de sa contribution à la prédiction. Réputé pour sa constance et sa précision locale, l'on y a souvent recours pour les données tabulaires, même si son calcul peut se révéler fastidieux.
LIME (Local Interpretable Model-Agnostic Explanations)	LIME crée des modèles linéaires locaux afin d'expliquer les prédictions pour des points de données particuliers, ce qui est utile notamment pour les modèles de classification. L'approche est modélisante, autrement dit LIME peut expliquer tout modèle ML, potentiellement toutefois au détriment d'une non-linéarité du modèle initial.
What-if Tool	Développé par Google, cet outil permet d'explorer visuellement les modèles de ML en offrant aux utilisatrices et utilisateurs la possibilité de simuler des scénarios hypothétiques et d'en comprendre l'impact sur les prédictions.
AIX360 (AI Explainability 360)	Une suite étendue d'algorithmes et de méthodes pour l'explicabilité et l'interprétabilité d'IBM Research censée couvrir différents besoins d'explication.
Phases de cycle de vie d'un système d'IA	La définition des sept phases du cycle de vie d'un système d'IA, telle que stipulée par la norme ISO/IEC22989, comprend l'initialisation, la conception, la vérification, l'implémentation, l'exploitation, la réévaluation et la mise hors service, afin de gérer les phases et les responsabilités tout au long du processus.
Explainable AI (XAI)	Technologies capables de présenter la prise de décision d'un système d'IA dans un langage simple, souvent avec le concours de méthodes de visualisation et d'explication
Skater	Un framework unifié encadrant l'interprétation des prédictions de modèle et l'explication du mode de fonctionnement des modèles complexes, qui simplifie la visualisation des influences et des importances des Features.
Traceability	La traçabilité des décisions et des processus, qui permet de remonter à l'origine et de suivre les étapes de traitement des données et des modèles.

4 Principes concernant l'IA en Suisse

4.1 «Directives d'IA pour l'administration fédérale»

Un groupe de travail interdépartemental, piloté par le Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI), a rédigé en 2019 un rapport sur les «Défis de l'intelligence artificielle» et, partant de là, a élaboré en collaboration avec le DETEC, les lignes directrices «Intelligence artificielle» pour l'administration fédérale.

Les lignes directrices reposent sur l'ordre juridique national et international en vigueur en Suisse, en particulier la Constitution fédérale de la Confédération suisse (Cst., RS 101) et les normes de la Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH), ainsi que sur d'autres principes juridiques fondamentaux pour le traitement de l'IA.

Ces lignes directrices constituent pour l'administration fédérale et les responsables des tâches administratives de l'État fédéral un cadre d'orientation global et doivent être respectées de manière spécifique dans les contextes suivants:

- lors de l'élaboration de stratégies sectorielles en matière d'IA;
- lors de l'introduction ou de l'adaptation de réglementations sectorielles spécifiques;
- lors du développement et de l'utilisation de systèmes d'IA au sein de l'administration fédérale;
- lors de contributions à l'élaboration d'un cadre réglementaire international en matière d'IA.

En 2025, le Conseil fédéral a par ailleurs décidé de ratifier la Convention sur l'IA du Conseil de l'Europe et de la transcrire dans le droit suisse. La convention engage en premier lieu les acteurs de niveau étatique et vise des principes fondamentaux tels que la transparence, la protection des données, la non-discrimination et la surveillance.

Les principaux éléments de la Stratégie suisse de réglementation sont les suivants:

- Mise en œuvre sectorielle: Les adaptations législatives se font de manière ciblée par secteur (par. ex. santé, transports), tandis que les thématiques touchant aux droits fondamentaux (par. ex. protection des données) font l'objet de réglementations plus générales.
- Approche réglementaire combinée: Outre les normes juridiquement contraignantes, des instruments non contraignants sont également développés, par ex. des déclarations volontaires et des normes sectorielles.

Trois objectifs stratégiques:

- Renforcement de la Suisse en tant que place économique et pôle d'innovation
- Sécurisation de la protection des droits fondamentaux et de la liberté économique
- Instauration de la confiance de la population dans une utilisation responsable de l'IA

4.2 Code de bonnes pratiques de la Confédération en matière de science des données et d'espaces de données dignes de confiance

Autre référence: le code de bonnes pratiques de la Confédération en matière de science des données centrée sur l'humain et digne de confiance. Les principes fondamentaux, tels que la protection des données et des informations, la sécurité des données, la non-discrimination, la traçabilité, la reproductibilité, l'objectivité, la sécurité des informations, la gouvernance des données, l'explicabilité, la transparence, la neutralité, le traitement éthique des données, et les résultats ont été consultés dans le cadre de la présente norme et pris en compte pour les thématiques correspondantes (OFS, 2023).

Le Code de bonnes pratiques en matière d'espaces de données dignes de confiance offre des recommandations ou un guide aux acteurs du secteur privé, des milieux universitaires et de la société civile, qui tous sont invités à y souscrire. Le code spécifie la conception d'espaces de données dignes de confiance à partir de quatre principes clés, à savoir la transparence, le contrôle, l'équité et l'efficacité. Chacun de ces quatre principes contient une série de recommandations pratiques assorties de mesures correspondantes pour une possible mise en œuvre. Les gains d'efficacité et le potentiel d'innovation découlant de l'échange de données profiteraient à la société dans sa globalité. Nous considérons par conséquent l'autodétermination numérique comme un objectif à long terme.

5 Phases du cycle de vie et rôles des Stakeholders des systèmes d'IA

Le modèle de cycle de vie des systèmes d'IA, tel qu'énoncé par la norme ISO/IEC22989, propose une description du développement d'un système d'IA, de son initialisation à sa mise hors service. Le modèle s'articule autour des sept phases du cycle de vie suivantes: initialisation, conception et développement, vérification et validation, implémentation, exploitation et surveillance, réévaluation et mise hors service (DIN SPEC 92001-3:2023-08).

Des processus et calendriers spécifiques peuvent apparaître au cours d'une ou de plusieurs phases du cycle de vie et certaines phases peuvent se répéter. Le modèle de cycle de vie doit être considéré non pas comme une recette, mais bien comme un instrument; divers modèles de cycle de vie sont possibles en fonction des différentes techniques de développement ou sphères d'application.

Se référant à la norme ISO/IEC22989, le présent document identifie douze rôles différents pour les Stakeholders, déclinés en six catégories: fournisseur d'IA, concepteur d'IA, client d'IA, partenaire d'IA, sujet d'IA et autorités concernées. Une même entité ou organisation peut assumer simultanément plusieurs rôles de Stakeholder et un seul rôle de Stakeholder peut être dévolu à plusieurs organisations.

Les exigences de la présente norme sont affectées aux différentes phases du cycle de vie des systèmes d'IA en vue de garantir une évaluation et une atténuation globales et continues des risques. Une telle précaution est nécessaire, car du développement au déploiement, à la maintenance et au démantèlement, les systèmes d'IA présentent des risques et des enjeux différents à chaque étape de leur cycle de vie. En ciblant les exigences sur chaque phase, il est possible de s'assurer que la protection des données, l'équité, la transparence, l'explicabilité et l'obligation de rendre des comptes sont prises en compte à toutes les étapes. On évite ainsi que les risques critiques ne soient identifiés que lors d'une phase ultérieure, l'implémentation des systèmes d'IA devenant ainsi plus sûre, plus éthique et plus digne de confiance.

Les Stakeholders à différents niveaux peuvent, le cas échéant, interagir avec les systèmes d'IA de l'une des quatre manières caractéristiques suivantes: développer, mettre à disposition, utiliser ou évaluer. Chaque type d'interaction caractéristique implique des besoins spécifiques en termes d'information.												
	Rôles des Stakeholders											
Phases de cycle de vie	**Fournisseur d'IA**		**Concepteur d'IA**	**Client d'IA**	**Partenaire d'IA**				**Sujet d'IA**		**Autorité pertinente**	
	Fournisseur de plateforme d'IA	Fournisseur de produit/de service d'IA	Développeur d'IA	Utilisateur d'IA	Intégrateur de systèmes d'IA	Fournisseur de données	Évaluateur d'IA	Auditeur d'IA	Sujet de données d'IA	Autres sujets	Auteur de policy	Régulateur
Initialisation	-	développer	développer	-	développer	-	-	-	-	-	-	-
Conception & développement	développer	développer	développer	-	développer	développer	-	-	-	-	-	-
Vérification & validation	-	développer, évaluer	développer, évaluer	-	développer, évaluer	-	évaluer	évaluer	-	-	évaluer	évaluer
Implémentation	développer	fournir, évaluer	développer, évaluer	fournir, utiliser	fournir, évaluer	-	-	-	-	-	-	-
Exploitation & surveillance	fournir	fournir, évaluer	développer, évaluer	fournir, utiliser	fournir, évaluer	-	évaluer	évaluer	utiliser	utiliser	-	-
Réévaluation	-	évaluer	-	-	-	-	évaluer	évaluer	-	-	évaluer	évaluer
Mise hors service	-	développer, évaluer	développer, évaluer	-	-	-	-	-	-	-	-	-

Tableau 1: Rôles de Stakeholders et phases de cycle de vie: Propre représentation (traduite) selon DIN SPEC 92001-3:2023-08, 2023)

6 Vue d'ensemble des normes et standards internationaux

L'élaboration des normes eCH sur l'IA a notamment impliqué la consultation des normes internationales. Comme cela est le cas pour les autres normes eCH, les normes de l'association se démarquent de leurs pendants internationaux en raison des lois de la Suisse et des besoins propres au pays. Il est important de préciser que les normes eCH dans le domaine de l'intelligence artificielle s'appuient sur des normes de référence internationales en termes de contenu et de méthodologie, dont les normes ISO/IEC JTC 1/SC 42.

Le graphique suivant présente le paysage international et européen de la normalisation dans le domaine de l'intelligence artificielle. Il illustre les acteurs centraux au niveau mondial (ISO, CEI, UIT) et européen (CEN, CENELEC, ETSI) ainsi que les comités techniques pertinents pour l'IA, notamment ISO/IEC JTC 1/SC 42 au niveau international et CEN-CENELEC JTC AI au niveau européen, qui assurent la coordination technique des normes d'IA.

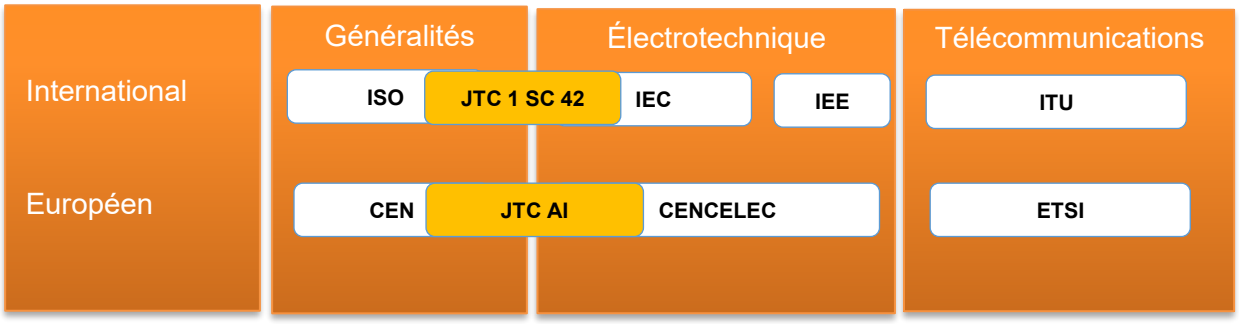

Figure 3: Vue d'ensemble des normes internationales

La présente norme eCH-0272 est en particulier harmonisée avec:

- ISO/IEC 22989:2022 (concepts et terminologie relatifs à l'IA),
- ISO/IEC 23053:2022 (cycles de vie des systèmes d'IA),
- ISO/IEC 24029-1:2021 (robustesse),
- ainsi que CEN-CENELEC JTC 21 Roadmap pour la mise en œuvre de l'AI Act de l'UE.

Les normes eCH optent à dessein pour une application contextuelle de ces règles internationales, non sans tenir compte du cadre juridique suisse (par. ex. Loi sur la protection des données, Cst., droit d'auteur).

Le tableau ci-dessous montre, à titre d'exemple, en quoi les contenus centraux de la norme eCH-0272 correspondent aux normes internationales pertinentes. Il se veut un guide pour les professionnels et les organisations souhaitant appliquer les normes eCH dans un contexte international ou les combiner avec des exigences conformes à la norme ISO/CEI.

Domaine / thème	Norme(e) ISO/IEC pertinente(s)	Correspondance dans eCH-0272	Compléments/spécificités suisses
Concepts et terminologie	ISO/IEC 22989:2022	Chapitre 3 (Terminologie)	Définition claire «Système d'IA» selon l'OCDE
Cycle de vie d'un système d'IA	ISO/IEC 23053:2022	Chapitre 4	Référence au contexte d'administration suisse et modèles de rôle
Transparence et explicabilité	ISO/IEC 24027 / ISO/IEC 62577 (in Draft)	Chapitre 7.8 & 7.9	Distinction entre objectifs d'acceptation/de protection
Obligation de rendre des comptes	DIN SPEC 92001-3:2023-08	Chapitre 7.7	Ancrage dans le cadre légal suisse
Protection des données	-	Chapitre 2.1, 7.3	Référence à la LPD au lieu du RGPD
Durabilité & équité	ISO/IEC 24030 (en préparation)	Chapitre 7.6, 7.10	Référence au Code de bonnes pratiques en matière de science des données de la Confédération

Tableau 2: Guide d'orientation des normes internationales et des normes eCH

Il est conseillé aux entreprises, en particulier celles ayant des activités internationales, d'appliquer les normes eCH en associant avec les normes ISO/CEI pertinentes. Tandis que la norme eCH-0272 est axée spécifiquement sur le cadre juridique et éthique suisse, les normes ISO/IEC font figure de référence mondiale pour la mise en œuvre technique.

- Une application hybride des deux familles de normes a pour avantages:
- le respect des règles suisses (par ex. LPD, Cst.),
- la possibilité de rattachement aux chaînes d'approvisionnement internationales,
- et une gouvernance cohérente et basée sur les risques pour les systèmes d'IA.

Il convient à cet égard de mentionner tout particulièrement la norme ISO/CEI 42001:2023, toute première norme internationale pour les systèmes de gestion de l'intelligence artificielle. Elle propose aux organisations un cadre structuré pour la gestion des risques et l'amélioration continue dans l'utilisation de systèmes d'intelligence artificielle – notamment en ce qui concerne la transparence, la responsabilité et les exigences de sécurité. La combinaison des normes ISO/IEC 42001 et eCH-0272 permet une mise en œuvre précise et contextualisée de la gouvernance de l'IA en Suisse.

7 Objectifs de protection et d'acceptation

Les objectifs de protection et d'acceptation suivants découlent de la discussion des chapitres 2 et 4. Ils traitent des risques et impacts potentiels que peut avoir l'utilisation de systèmes d'IA dans différents domaines. Le tableau suivant offre un vue d'ensemble centrale et les objectifs de protection et d'acceptation posent un cadre pour la présente norme. Ils rendent possible une description systématique des exigences et des mesures. Le respect des exigences contribue à la conformité des technologies d'IA avec les exigences éthiques, juridiques et sociétales. Il en résulte une base pour une utilisation sûre et responsable de l'IA dans différents domaines d'application.

Objectifs de protection Objectifs visant à garantir la sécurité, la protection des données et les droits des individus. Ces objectifs ont pour priorité de réduire au minimum les risques pouvant découler de l'utilisation de l'IA, tels que les atteintes à la protection des données, la discrimination ou les mauvaises décisions. Les objectifs de protection sont fortement axés sur les règles juridiques et les normes éthiques en vigueur afin de s'épargner les conséquences négatives des technologies d'IA.

Objectifs d'acceptation Objectifs visant à promouvoir la confiance et l'acceptation des systèmes d'IA au sein de la société. Il s'agit notamment de mesures visant à améliorer la transparence, l'explicabilité et la perception positive de l'IA par les utilisatrices et utilisateurs et les Stakeholders. Les objectifs d'acceptation incluent également la promotion d'un environnement au sein duquel l'IA est considérée comme un ingrédient sûr et précieux de l'innovation et du progrès technologique.

Objectifs de protection

Objectif	**Description**	**Analogies avec les normes et corpus de règles suivants**
Généralités	Cet objectif de protection inclut toutes les exigences qui ne sont pas couvertes par les autres objectifs de protection, tels que les droits fondamentaux, la protection des données, les droits d'auteur, etc., mais qui peuvent pourtant avoir des conséquences négatives pour les individus, les organisations ou la société. Cela comprend également des exigences dans des domaines tels que l'innovation, les progrès technologiques ou les opportunités de marché.	-
Droits fondamentaux	Cet objectif de protection se réfère au respect et à la protection des droits humains fondamentaux tels qu'ils sont inscrits dans les lois nationales et internationales. Les systèmes d'IA ne sont pas autorisés à prendre des décisions discriminatoires ou injustes qui portent atteinte aux droits à la vie, à la liberté, à la dignité ou à l'égalité.	Constitution fédérale, convention sur l'IA, AI-Act

Objectif	Description	Analogies avec les normes et corpus de règles suivants
Protection des données	Cet objectif de protection concerne la protection de la sphère privée et des données à caractère personnel des utilisatrices ou utilisateurs. Cela concerne la collecte, le traitement et le stockage des données lors de l'utilisation de systèmes d'IA. Le droit des utilisatrices et utilisateurs à décider de ce qu'il advient des informations les concernant doit être préservé.	Loi sur la protection des données
Droits d'auteur	Il s'agit là de la protection de la propriété intellectuelle et du respect des droits d'auteur lors de l'utilisation d'œuvres ou de données traitées ou générées par l'IA.	Loi sur le droit d'auteur
Autonomie	Cet objectif de protection porte sur la capacité des utilisatrices et utilisateurs à prendre des décisions de manière indépendante et en toute connaissance de cause. Les systèmes d'IA devraient respecter l'autonomie des personnes, ne pas les manipuler et ne pas prendre de décisions qui les privent de leur droit à choisir librement.	Lignes directrices IA administration fédérale,
Équité	Il est ici question de l'égalité de traitement de toutes les personnes et groupes concernés. Les systèmes d'IA ne doivent favoriser aucune discrimination fondée sur le sexe, l'origine ethnique, la religion ou d'autres caractéristiques.	Code de bonnes pratiques en matière de science des données de la Confédération
Obligation de rendre des comptes	Cet objectif impose que les responsables des systèmes d'IA puissent répondre de tout comportement fautif ou dommage. Il doit être clairement établi à qui incombe la responsabilité des décisions prises par l'IA.	Lignes directrices IA administration fédérale,

Tableau 3: Objectifs de **protection**

Objectifs d'acceptation

Objectif	Description	Analogies avec les normes et corpus de règles suivants
Transparence	La transparence impose une présentation, qui soit compréhensible et puisse être tracée, du fonctionnement des systèmes d'IA. Les utilisateurs et les décideurs doivent être en mesure de comprendre les méthodes, les décisions et les effets des systèmes d'IA.	Code de bonnes pratiques en matière de science des données de la Confédération
Explicabilité	Cet objectif exige que les décisions et les actions d'un système d'IA puissent être expliquées de façon compréhensible pour les humains.	Lignes directrices IA administration fédérale,

Tableau 4: Objectifs d'acceptation

Évaluation des exigences de «notation» et de «classification»

Dans la conception des systèmes d'IA, l'évaluation systématique des exigences joue un rôle crucial pour répondre aux normes légales et éthiques et pour renforcer la confiance des utilisatrices et utilisateurs. L'évaluation des exigences dans ce chapitre s'appuie sur les paramètres suivants.

Notation comme degré d'obligation: La notation («IMPÉRATIF», «RECOMMANDÉ», «FACULTATIF») indique dans quelle mesure une exigence doit effectivement être mise en œuvre. Cette notation sert d'instructions d'action, qui indiquent aux organisations dans quelle mesure elles doivent suivre rigoureusement les règles.

Classification comme estimation des risques: La classification («élevé», «moyen», «faible») illustre le niveau de risque et fournit un contexte de pertinence et l'urgence relatif à l'exigence. La classification aide les organisations à identifier les exigences particulièrement critiques pour la protection des données, l'équité ou l'explicabilité de leurs systèmes d'IA.

8 Exigences relatives aux systèmes d'IA

La matrice ci-dessous vient compléter de manière pratique les exigences énoncées par les tableaux 8.1 à 8.9 et aide ainsi les organisations à sélectionner de façon ciblée les mesures pertinentes selon le système d'IA, du domaine d'application et du profil de risque.

Alors que les tableaux 8.1 à 8.9 détaillent les exigences en fonction des objectifs de protection et d'acceptation, des phases du cycle de vie et des notations, cette matrice offre une orientation simplifiée du point de vue de l'utilisateur ou de l'utilisatrice. Elle permet d'effectuer une première attribution des exigences à respecter en priorité pour certaines combinaisons de type d'IA, de contexte d'utilisation et de risque.

La matrice n'a pas vocation à se substituer aux tableaux existants, mais vient les compléter dans le but de faciliter leur mise en œuvre dans la pratique. Elle sera développée plus avant dans les versions à venir de la norme et assortie d'exemples d'application caractéristiques.

Type d'IA	Contexte d'utilisation	Classe de risque	Exigences d'objectif de protection	Objectifs d'acceptation concernés	Exigences d'objectif d'acceptation	Objectifs d'acceptation concernés
Système symbolique	Administration interne	Basse	A3, A16, A18	Obligation de rendre des comptes	A22, A23	Transparence, explicabilité
Système basé sur le ML	Administration avec effets extérieurs	Élevée	A1, A2, A3, A4, A5, A6, A7, A8, A14, A15, A16, A17, A18, A19, A20	Protection des données, équité, autonomie, droits fondamentaux, obligation de rendre des comptes	A21, A24, A25, A26, A27	Explicabilité, transparence
IA générative	Interaction citoyenne	Moyenne	A2, A6, A14, A15, A16, A19	Protection des données, équité, autonomie	A21, A23, A24, A25	Transparence, explicabilité
Système statistique déclaratif	Utilisation technique interne	Moyenne	A3, A6, A10, A12, A16, A18	Obligation de rendre des comptes, droits fondamentaux	A22, A23	Transparence, explicabilité
Système hybride	Décisions de prestation	Élevée	A1, A2, A3, A4, A5, A6, A7, A8, A14, A15, A16, A17, A18, A19, A20	Protection des données, équité, droits fondamentaux, obligation de rendre des comptes	A21, A24, A25, A26, A27	Explicabilité, transparence

Tableau 5: Matrice d'orientation, tableaux des exigences

Les exigences suivantes s'inspirent d'une approche fondée sur les risques. Le contenu des exigences s'appuie sur le chapitre 2 «Bases légales» et le chapitre 4 « Principes de l'IA en Suisse».

La colonne Cycle de vie de l'IA spécifie les cycles de vie pertinents pour les exigences, conformément au chapitre 5 Phases du cycle de vie et rôles des Stakeholders des systèmes d'IA. On distingue les phases suivantes: Initialisation, conception & développement, vérification & validation, implémentation, exploitation & surveillance, réévaluation, mise hors service

Les destinataires de chacune des exigences dans les sous-chapitres suivants sont identifiés par l'attribut «Cycle de vie de l'IA». Se reporter au chapitre 5 «Life Cycle et rôles des Stakeholders systèmes d'IA» pour en savoir plus sur l'attribution des destinataires à chaque cycle de vie de l'IA.

Les exigences et mesures décrites dans les chapitres suivants ne sont en outre pas exhaustives. D'autres exigences ou mesures peuvent être adoptées en fonction de l'application et du contexte du système d'IA.

Informations complémentaires: Concernant les exigences et les mesures relatives aux objectifs de protection et d'acceptation mentionnés, des informations complémentaires relatives à la traçabilité sont mentionnées à chaque fois à la fin d'un sous-chapitre en vue de garantir cette traçabilité. Cela permet d'indiquer les principes dont s'inspirent les exigences. Pour chacune des exigences, une ou plusieurs sources ont été consultées. Les paramètres sont affichés à titre d'exemple avec le tableau suivant.

Description			**Traceability**		
			Explicite		Implicite
N°	**Notation**	**Exigence (minimale)**	**Loi**	**Étude/ rapport/norme int.**	**E-Input**
A1	IMPÉRATIF	Lors du traitement de données à caractère personnel par un système d'IA, les personnes concernées doivent être avisées de la finalité, de l'identité et des coordonnées du ou de la responsable du traitement et, le cas échéant, des destinataires de ces données à caractère personnel.	LPD	n/a	Oui/Non?
A2	FACULTATIF	…….	n/a	n/a	Oui

Tableau 6: Légende Traceability

La numérotation sert à l'identification des différentes exigences (A1, A2, etc.) et des mesures correspondantes (A1M1, A1M2, etc.). La notation renvoie aux règles du chapitre 1.5, généralement utilisées dans les normes eCH. Les exigences sont saisies individuellement.

8.1 Généralités

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A1	Conception & développement, Vérification & validation	IMPÉRATIF	Le champ d'application du système d'IA est pris en compte pour l'évaluation de la classification des risques. Justification: Le champ d'application peut avoir des répercussions considérable sur la façon dont un système d'IA peut générer des risques. Un système d'IA employé dans un domaine critique, comme la médecine, présente potentiellement davantage de risques qu'un système employé dans un autre moins critique. Les risques associés aux systèmes d'IA ne proviennent pas de l'IA elle-même en tant qu'entité autonome, mais des décisions prises par l'humain au moment du développement, du training, de la mise en œuvre et de l'exploitation de ces systèmes. Les systèmes d'IA constituent des outils dont les effets dépendent directement des données, des algorithmes et des contextes d'utilisation sous-jacents.	**A1M1: Documenter le champ d'application:** Élaboration d'une documentation détaillée, qui décrit clairement le champ d'application prévu pour le système d'IA. Les scénarios dans lesquels le système est utilisé, le groupe cible et les tâches spécifiques qu'accomplit le système d'IA en font partie. La documentation devrait etre rédigée dans un langage simple. **A1M2: Identifier les répercussions négatives:** Identification des répercussions négatives potentiels dans ce domaine. **A1M3: Comparer les solutions** Comparaison avec les solutions existantes (pour autant qu'il y en ait). **A1M4: Publier la documentation du champ d'application** La documentation du champ d'application est rendue accessible au public.	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A2	Conception & développement, Vérification & validation	IMPÉRATIF	La ou les méthodes d'apprentissage utilisées par le système d'IA, le modèle d'intelligence artificielle, les données de formation et le processus de formation sont enregistrés et évalués à des fins d'évaluation de la classification des risques. La méthode d'apprentissage et les données sous-jacentes peuvent conduire à des erreurs et des biais masqués Quelques méthodes sont plus sensibles aux surajustements ou imposent une vérification plus attentive que d'autres. De plus, un tel manque de transparence peut déboucher sur une classification des risques plus élevée, en particulier lorsque le système d'IA est mis à contribution dans des domaines d'application critiques où la traçabilité des décisions figure parmi les exigences.	**A2M1: Déclarer la/les méthode(s) d'apprentissage utilisée(s)** La méthode d'apprentissage, comme l'apprentissage supervisé, non supervisé ou renforcé, est consignée dans la documentation des modèles. Le pourquoi de l'adaptation de la méthode au problème est également décrit. La ou les méthodes d'apprentissage devraient être décrites de la manière la plus compréhensible qui soit. **A2M2: Déclarer le modèle d'IA:** Le type de modèle d'IA (par ex. arbre de décision, réseau neuronal, transformateur) est exposé en détail, y compris le nombre de couches ou de paramètres pour les modèles complexes. Ce manque de transparence peut déboucher sur une classification des risques plus élevée, en particulier lorsque le système d'IA est mis à contribution dans des domaines d'application critiques où la traçabilité des décisions figure parmi les exigences. **A2M3: Documenter les données et le processus de training:** Les ensembles de données de training utilisés sont nommés et, si possible, rendus accessibles. Le processus de formation est décrit.	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				A2M4: Analyser les sources de données et de leur qualité. **A2M5:** Évaluation des questions de biais et d'équité.	
A3	**Vérification & validation** Implémentation Exploitation & surveillance Réévaluation	IMPÉRATIF	Le comportement du système d'IA et les décisions qu'il prend s'accompagnent de contrôles en continu afin de s'assurer qu'ils sont en phase avec les fonctions prévues et les exigences éthiques. Les conséquences sociales et éthiques de l'utilisation d'un système d'IA doivent également être prises en compte pour évaluer sa classification des risques.	**A3M1: Décrire le Worst Case Scenario** Description des worst-case scenarios (pires scénarios) possibles. **A3M2: Plans d'urgence:** Plans d'urgence et mesures d'atténuation des risques. **A3M3: Introduire les processus:** Mise en place de processus servant à identifier et à traiter les conséquences sociales et éthiques de l'utilisation des systèmes d'IA. **A3M4: Implémenter et évaluer les protocoles d'audit:** Mise en place de protocoles d'audit et de processus de surveillance approfondis visant à garantir que le comportement du système d'IA et sa prise de décision font l'objet d'un examen continu de leur conformité avec les fonctions prévues et les exigences éthiques. Ces protocoles peuvent contribuer à identifier et à traiter les biais, les comportements fautifs ou d'autres facteurs de risque, ce qui	

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				accroît la sécurité et la crédibilité du système.	
A4	Conception & développement, Vérification & validation, Implémentation	IMPÉRATIF	L'évaluation des risques d'un système d'IA devrait impliquer de collecter des valeurs mesurables. Les valeurs suivantes au minimum devraient être collectées: valeur pour la classification des risques ainsi que valeur pour le champ d'application.	**A4M1: Classification des risques:** Quantification du risque global sur la base des exigences précédentes. Champ d'application: Mesures tirées d'A1 (30%) - Méthode d'apprentissage: Mesures tirées d'A2 (20%) - Valeur de risque: Mesures tirées d'A3 (50%) Justification: La mesure et la quantification directes du risque fournissent une indication évidente des dangers potentiels inhérents au système. La pondération de 50% indique clairement que le profil de risque direct du système d'IA est des plus importants. Les pourcentages indiqués illustrent l'importance relative de chaque dimension. Il a beau être important, le champ d'application n'en est pas moins éclipsé par la valeur de risque réelle du système, donc seulement 30%. La méthode d'apprentissage est certes critique, mais pas autant que les deux autres catégories, d'où 20%. La valeur de risque, qui est l'indicateur direct et global du danger du système, affiche en revanche le pourcentage le plus élevé, soit 50%.	Basse

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A5		IMPÉRATIF	L'utilisation des systèmes d'IA est responsable et durable. Les impacts des systèmes d'IA sur leur durabilité environnementale, en particulier sur l'efficacité énergétique de la programmation et des techniques mises en œuvre, doivent être évaluées et minimisées.	**A5M1: Durabilité écologique:** Les impacts écologiques (consommation de ressources lors du développement et de l'utilisation) devraient être analysés de manière fiable et publiée. **A5M2: Durabilité économique:** La durabilité économique (concentration du marché, dépendance vis-à-vis des fournisseurs de technologies dominant le marché, etc.) doit être analysée de manière fiable et publiée. **A5M3: Durabilité sociale:** La durabilité sociale (évaluation et responsabilité des conditions de travail dans la chaîne d'approvisionnement de l'IA) doit être analysée de manière fiable et publiée. **A5M4:** Contrôle et mesures: Mise en place de processus servant à déterminer et à optimiser la durabilité écologique de l'utilisation des systèmes d'IA.	

Tableau 7: Exigence systèmes d'IA – Généralités

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A1	-	Lignes directrices Intelligence artificielle Confédération, AI-Act	Équipe de base GS IA
A2	-	-	Équipe de base GS IA
A3	-	-	Équipe de base GS IA
A4	-	-	Équipe de base GS IA
A5			Consultation publique

Tableau 8: Exigence systèmes d'IA – Généralités informations complémentaires

8.2 Droits fondamentaux

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A5	Conception & développement, Exploitation & surveillance, Réévaluation	IMPÉRATIF	La dignité humaine doit être respectée au moment de développer et d'utiliser des systèmes d'IA. Les systèmes d'IA doivent être conçus et déployés avec pour objectif prioritaire la promotion du bien-être des personnes. Ils doivent cibler les besoins actuels des personnes, tenir compte de leurs contraintes et de leurs points forts et les aider dans l'accomplissement de leurs tâches.	**A5M1: Procéder à des analyses des besoins:** Réalisation d'analyses approfondies des besoins en vue de comprendre les besoins actuels et à venir des groupes cibles. **A5M2: Obtenir des retours** Programmation de dialogues réguliers et sollicitation de réactions des Stakeholders pertinents, dont les utilisateurs finaux, les personnes concernées, les experts techniques et la société civile. Mise en place de processus visant une prise en compte adéquate des retours d'information dans la révision des systèmes d'IA. **A5M3: Accessibilité et convivialité:** Veiller à ce que les systèmes et applications d'IA soient conçus de manière accessible et conviviale pour en faciliter l'accès et l'utilisation à grande échelle. (eCH-0059). Mise à disposition, dans la mesure du possible, d'une offre alternative non numérique. **A5M4: Usability-Tests:** Réalisation d'Usability Tests auprès d'utilisateurs issus de différents groupes cibles à des fins d'amélioration continue de la convivialité. Mise en place de processus visant une prise en compte adéquate des résultats de ces tests dans la révision des systèmes d'IA.	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A6	Initialisation, conception & développement, exploitation & surveillance	IMPÉRATIF	Les outputs d'un système d'IA ne doivent pas porter atteinte aux droits de la personnalité des individus. Les restrictions des droits fondamentaux ne sont pas toujours manifestes.	**A6M1: Processus de contrôle:** Mise en place de processus de contrôle techniques et organisationnels censés éviter toute atteinte aux droits de la personnalité. **A6M2: Analyses d'impact:** Réalisation analyse d'impact visant à évaluer les conséquences sur les personnes en termes de droits fondamentaux. **A6M3: Audits:** Réalisation régulière d'audits par des instances de contrôle internes ou externes.	Élevée

Tableau 9: Exigence systèmes d'IA - Droits fondamentaux

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A5	Constitution fédérale	Lignes directrices intelligence artificielle, Code de bonnes pratiques en matière de science des données de la Confédération	-
A6	-	Lignes directrices intelligence artificielle, Code de bonnes pratiques en matière de science des données de la Confédération Convention sur l'IA du Conseil de l'Europe	

Tableau 10: Exigence systèmes d'IA – Droits fondamentaux informations complémentaires

8.3 Protection des données

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A7	Initialisation, Conception & développement, Vérification & validation, Implémentation, Exploitation & surveillance, Réévaluation Mise hors service	IMPÉRATIF	Les principes de traitement des données sont respectés lors de l'utilisation de systèmes d'intelligence artificielle (par ex. la LPD, les lois cantonales, etc.).	**A7M1: Principes de traitement des données:** Respect des principes fondamentaux de traitement des données lorsque des données à caractère personnel sont traitées par un système d'IA ou servent aussi de base aux modèles (données de formation). **A7M2: Informer les personnes concernées:** Information des personnes concernées à propos du traitement des données par un système d'IA et de la prise de décision individuelle automatisée.	Élevée
A8	Conception & développement, exploitation & surveillance	IMPÉRATIF	Il faut veiller au respect des droits fondamentaux lorsque des données à caractère personnel particulièrement sensibles sont traitées à grande échelle ou lorsque de vastes zones publiques font l'objet d'une surveillance systématique. Le risque est particulièrement élevé lorsque des données à caractère personnel sensibles sont traitées à grande échelle ou lorsque de vastes zones publiques font l'objet d'une surveillance systématique.	**A8M1: Analyse d'impact sur la protection des données:** Réalisation préalable (avant utilisation) d'une analyse d'impact relative à la protection des données lorsqu'un traitement prévu (par ex. un système d'IA est utilisé pour le traitement de données à caractère personnel) présente un risque important pour la personnalité ou les droits fondamentaux. Remarque: une analyse d'impact relative à la protection des données est notamment requise dans les cas de figure suivants: - évaluation systématique et approfondie des aspects personnels des personnes physiques, y compris le profilage;	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				- traitement à grande échelle de données sensibles; - surveillance systématique approfondie des domaines accessibles au public. **A8M2: Anonymisation:** Anonymisation des données à caractère personnel Il peut être possible de tirer implicitement des conclusions sur des personnes sans recourir aux données à caractère personnel. Le risque d'inférence doit également être pris en compte en cas de traitement de données «anonymisées». **A8M3: Formations** Formation du personnel faisant usage du système d'IA.	

Tableau 11: Exigence systèmes d'IA – Protection des données

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A7	Loi sur la protection des données Article 6 LPD Principes du traitement des données, notamment les principes de licéité, de transparence, de proportionnalité et de finalité du traitement des données à caractère personnel. Article 19 LPD Devoir d'informer lors de la collecte de données à caractère personnel Article 21 LPD Devoir d'informer en cas de décision individuelle automatisée	-	-
A8	Loi sur la protection des données Article 22 LPD Analyse d'impact relative à la protection des données	-	-

Tableau 12: Exigence systèmes d'IA – Protection des données informations complémentaires

8.4 Droits d'auteur

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A9	Initialisation, conception & développement, exploitation & surveillance	IMPÉRATIF	Les systèmes d'IA doivent respecter la propriété intellectuelle. Concernant l'utilisation de contenus par des systèmes d'IA, la Suisse n'a encore défini aucune évaluation juridiquement contraignante des droits d'auteur. On peut certes imaginer que le training de systèmes d'IA et l'élaboration d'outputs d'IA donnent lieu à des utilisations d'œuvres soumises au droit d'auteur. Cependant, il n'existe pas encore d'évaluation contraignante à ce sujet.	**A9M1: Processus:** Développement et mise en œuvre d'un processus qui garantit que tous les contenus traités par un système d'IA n'enfreignent pas le droit d'auteur avant d'être utilisés ou reproduits. **A9M2: Audits:** Réalisation d'audits réguliers des systèmes d'IA visant à s'assurer que la loi sur le droit d'auteur est bien respectée lors de l'utilisation des données de training et des outputs du système d'IA. Des mesures correctives doivent être mises en place, si besoin est. **A9M3: Gestion des licences:** Mise en place d'un système de gestion et de surveillance des licences, qui contrôle et documente l'utilisation des produits faisant l'objet de protections. **A9M4: Systèmes de filtrage de contenu:** Utilisation de systèmes de filtrage de contenu (content filtering) capables de détecter et de bloquer automatiquement le matériel protégé par des droits d'auteur. **A9M5: Mesures:**	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				Intégration de mesures techniques de protection empêchant la copie et la diffusion illégale de matériel protégé par des droits d'auteur. Remarque: les mesures techniques ne doivent pas être les seules utilisées. **A9M6: Obtenir des licences:** Si des œuvres soumises au droit d'auteur sont utilisées pour le training de systèmes d'IA et pour la création d'inputs d'IA, il convient de vérifier si, dans un cas concret, une barrière du droit d'auteur peut s'appliquer (comme la barrière scientifique ou le droit de citation par exemple). Dans le cas contraire, des licences individuelles devraient être obtenues auprès des titulaires des droits.	

Tableau 13: Exigence systèmes d'IA – Droits d'auteur

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A9	Droits d'auteur Article 2 LDA Définition de ce que le droit d'auteur entend comme œuvre. Article 10 LDA Utilisation de l'œuvre	-	-

Tableau 14: Exigence systèmes d'IA – Droits d'auteur informations complémentaires

8.5 Autonomie

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A10	Conception & développement, exploitation & surveillance	IMPÉRATIF	Les systèmes d'IA doivent faire l'objet d'une évaluation continue des risques relatifs à l'autonomie des personnes concernées.	**A10M1: Documentation** Réalisation de documentation et analyses régulières des systèmes d'IA et examen de mesures supplémentaires (par ex. publication des résultats). Cette analyse devrait être effectuée en continu tout au long du cycle de vie du système d'IA et prendre en compte les risques identifiés. Dans certains cas, la documentation seule peut suffire.	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A11	Conception & développement, exploitation & surveillance	IMPÉRATIF	**Protection des informations** Les systèmes d'IA doivent protéger les informations, telles que les secrets commerciaux, etc. de tout abus. **Sécurité des informations** L'intégrité, la disponibilité, la confidentialité et l'authenticité des données sont assurées au moyen de mesures techniques et organisationnelles.	**A11M1: Audits réguliers** Réalisation régulière d'audits des systèmes d'IA visant à identifier et à contrer les risques. **A11M2: Conditions d'utilisation** Examen des conditions d'utilisation des systèmes d'IA. **A11M3: Concept SIPD** Le concept SIPD sert de base à la détermination de mesures de sécurité des informations, selon HERMES.**A11M4: Sécurité du système**Le concept d'accès et d'autorisation, les procédures d'authentification, les procédures de chiffrement, le back-up et les autres mesures techniques de sûreté de l'information reflètent l'état le plus récent des connaissances techniques	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A12	Exploitation & surveillance	IMPÉRATIF	Les processus et systèmes de surveillance doivent garantir la robustesse des systèmes d'IA.	**A12M1Systèmes de détection des anomalies** Développement et utilisation d'algorithmes de détection des anomalies à partir de l'output du système d'IA. **A12M2: Suivi en temps réel** Pour autant que le niveau de risque l'impose, mise en place de services de surveillance en temps réel chargés de contrôler en permanence l'activité du système et d'émettre immédiatement une alerte en cas de détection d'anomalies. **A12M3: Rapports et journalisation** Mise en œuvre de systèmes de journalisation qui consignent, avec force détails, toutes les activités dans l'optique d'une analyse ultérieure et de la préservation des preuves d'éventuels incidents. Remarque: outre les systèmes automatisés, il faut impérativement prendre des mesures organisationnelles servant à signaler manuellement tout comportement non conforme.	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A13	Conception & développement, exploitation & surveillance	IMPÉRATIF	Implication d'experts humains dans les processus de décision et de contrôle des systèmes d'IA	**A13M1: Utilisation de mécanismes Human-in-the-loop par exemple:** ▪ Surveillance et (si nécessaire) intervention dans les décisions du système d'IA ▪ Boucles de feedback afin d'améliorer en permanence les décisions et suggestions du système d'IA ▪ Un système d'IA redirige les cas complexes et incertains vers des experts humains ▪ Les processus de prise de décision sont conçus de manière à ce que l'humain et le système d'IA fassent valoir leurs atouts respectifs. ▪ L'expertise (technique) des experts humains est intégrée au training des systèmes d'IA. **A13M2: possibilité de contact de seuil faible:** Dans le cas de processus de décision automatisés, une possibilité de contact à bas seuil doit être mise en place pour les recours lorsque les personnes concernées ne sont pas d'accord avec la décision prise.	

Tableau 15: Exigence systèmes d'IA – Autonomie

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A10	-	Principe de l'OCDE	-
A11	-	Code de bonnes pratiques en matière de science des données de la Confédération	-
A12	-	Code de bonnes pratiques en matière de science des données de la Confédération	-
A13	-	-LPD art. 21	

Tableau 16: Exigence systèmes d'IA – Autonomie informations complémentaires

8.6 Équité

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A14	Initialisation Conception & développement, Vérification & validation, Implémentation, Exploitation & surveillance, Réévaluation Mise hors service	IMPÉRATIF	Les systèmes d'IA ne doivent pas défavoriser les personnes ou groupes de personnes sur la base de caractères discriminatoires précis. Selon l'art. 8 de la Constitution fédérale, cela concerne notamment la discrimination fondée sur le handicap, le sexe, la filiation, la langue, l'origine ethnique ou sociale, les caractéristiques génétiques, l'orientation sexuelle, la religion ou les convictions, sans oublier les opinions politiques ou autres d'une personne. Le système d'IA doit garantir le principe de l'égalité de traitement et faire en sorte de ne créer ni ne renforcer des discriminations illégales fondées sur ces caractéristiques au moment de de la prise de décision.	**A14M1: Documentation:** Documenter les décisions ayant trait à la conception, ainsi que les éventuels compromis et conflits en matière d'objectifs. **A14M2: Analyse d'impact:** Dans le cas de l'utilisation d'un système d'IA, une analyse d'impact est réalisée avant, durant et après l'utilisation du système d'IA, qui tient compte des effets de l'utilisation d'un tel système sur les droits fondamentaux. Celle-ci inclut les éléments suivants: ▪ une description des procédures des exploitants; ▪ une description de la période et de la fréquence pendant lesquelles le système d'IA doit être utilisé; ▪ les catégories de personnes physiques et de groupes de personnes susceptibles d'être concernées par son utilisation dans un contexte précis; ▪ les risques spécifiques de dommages, ▪ une description de la mise en œuvre des mesures de surveillance	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				humaine conformément aux instructions d'exploitation; ▪ les mesures à prendre lorsque ces risques surviennent, y compris les règles de gouvernance interne et les mécanismes de recours. **A14M3: Audits réguliers** Réalisation d'audits réguliers des systèmes d'IA afin de mettre en évidence les risques de discrimination et de prendre les mesures appropriées pour éviter le risque.	
A15	Conception & développement, exploitation & surveillance Réévaluation	RECOMMANDÉ	Les données de training des systèmes d'IA doivent être représentatives de l'intégralité des personnes concernées et pas uniquement de sous-groupes.	**A15M1: Collecte des données diversifiée** Extension active des jeux de données visant à englober une variété de groupes démographiques et de scénarios. Cela inclut l'acquisition et l'intégration ciblées de données issues de diverses sources, qui représentent différents sexes, groupes d'âge, milieux socio-économiques, régions géographiques et autres dimensions pertinentes en lien avec la diversité. **A15M2: Analyse et correction des biais** Réalisation régulière d'une analyse des biais destinée à identifier les biais existants dans les données de training. Il convient ensuite de prendre des mesures destinées à corriger ces biais, en ajustant les jeux de	Moyenne

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				données, en recourant par exemple à des méthodes de resampling ou en appliquant des techniques d'apprentissage automatique ayant pour but l'équité et l'égalité de traitement. Remarque : la liste des mesures n'est pas exhaustive. Des mesures supplémentaires peuvent être envisagées.	

Tableau 17: Exigence systèmes d'IA – Équité

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A14	Constitution fédérale article 8	Code de bonnes pratiques en matière de science des données de la Confédération	-
A15	-	Code de bonnes pratiques en matière de science des données de la Confédération	Équipe de base GS IA

Tableau 18: Exigence systèmes d'IA – Équité informations complémentaires

8.7 Obligation de rendre des comptes

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A16	Conception & développement, Vérification & validation, Implémentation, Exploitation & surveillance, Réévaluation Mise hors service	IMPÉRATIF	Les responsabilités concernant les systèmes d'IA Un rôle responsable (par ex. personne, service) doit être défini concernant le respect des exigences et des mesures.	**A16M1: Création d'un framework de responsabilité** Développement d'un framework détaillé qui décrit les responsabilités du système d'IA. Cela devrait couvrir les rôles et compétences en matière de sélection des données, de training du modèle, de validation, de mise à disposition et de maintenance en continu. Le framework devrait aussi assurer la conformité avec les normes pertinentes – ISO/IEC 22989 et DIN SPEC 92001-3:2023-08, etc. et prodiguer des conseils quant à la façon de documenter et de communiquer les responsabilités. **A16M2: Vérification et adaptation régulières des responsabilités** Mise en œuvre des processus qui permettent de vérifier et d'adapter régulièrement les responsabilités afin de réagir aux modifications dans le système d'IA ou dans les normes pertinentes. Ceci pourrait se faire par le biais d'un système de Compliance Management propre à garantir que la conformité du système d'IA et de son fonctionnement aux dispositions les plus récentes de la norme ISO/IEC 22989 et de la norme	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				DIN SPEC 92001-3:2023-08. Il est notamment question ici de former régulièrement les personnes impliquées afin de s'assurer qu'elles soient au fait des toutes dernières exigences et Best Practices en matière d'IA. **A16M3: Introduction d'un mécanisme de réclamation et d'appel concernant les décisions de l'IA.** Tout système d'IA qui prend ou appuie des décisions susceptibles d'affecter les droits fondamentaux des personnes concernées doit mettre à disposition un mécanisme formel d'appel et de recours. Cela permet aux personnes concernées, de façon individuelle ou collective (le cas échéant avec l'aide d'organisations représentatives), i) d'avoir accès à des informations et à des explications concernant les décisions prises et ii) d'introduire un recours lorsqu'elles jugent une décision injuste ou erronée.	
A17	Conception & développement, Vérification & validation, Implémentation, Exploitation & surveillance, Réévaluation Mise hors service	IMPÉRATIF	Le processus de training et les outputs des systèmes d'IA doivent être documentés à toutes les phases du cycle de vie afin de garantir la traçabilité ainsi que la transparence.	**A17M1: Historique des versions et documentation des jeux de données** Pour chaque phase de développement d'un système d'IA, les jeux de données utilisés, y compris leurs versions et les modifications apportées, doivent être intégralement documentés. Cela pourrait se faire par le biais d'un système de contrôle des versions qui consigne également l'origine des données	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				(Data Provenance) et les modifications apportées aux jeux de données. Il s'agit entre autres de consigner les méthodes de collecte, de correction, d'enrichissement et de partage des données. Dans un souci de transparence, ces informations devraient, dans la mesure du possible, être accessibles au public.**A17M2: Protocoles de décision et outils d'explicabilité** Un système de journalisation des décisions devrait être mis en place afin de consigner avec précision comment et pourquoi le système d'IA a pris certaines décisions. À cela peut s'ajouter l'utilisation d'outils d'explicabilité (Explainable AI, XAI) qui assurent la transparence des facteurs et de la logique que sous-tendent les prédictions du modèle. Parmi ces outils, il peut y avoir des représentations visuelles des décisions de modèle, des analyses de sensibilité ou des analyses de contribution des features. Ces informations devraient être mises à la disposition des personnes concernées par les décisions, qui en font la demande, selon une procédure simple. Dans un souci de transparence, ces informations devraient, dans la mesure du possible, être accessibles au public.	

Tableau 19: Exigence systèmes d'IA – Obligation de rendre des comptes

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A16	-	Norme internationale ISO et DIN SPEC	Équipe de base GS IA
A17	-	Lignes directrices Intelligence artificielle Confédération	-

Tableau 20: Exigence systèmes d'IA – Obligation de rendre des comptes informations complémentaires

8.8 Transparence

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A18	Conception & développement, Vérification & validation, Implémentation, Exploitation & surveillance, Réévaluation Mise hors service	IMPÉRATIF	Pour chaque système d'IA, un rapport de training du modèle est établi préalablement à la mise en œuvre. Le contenu décrit dans les mesures doit figurer dans ce rapport. Ce rapport est rendu public dans un souci de transparence. Le niveau de détail du rapport peut varier en fonction du contexte propre au système d'IA.	**A8M1: Rapport de training du modèle** **Introduction et objectif** Une description succincte du problème à résoudre et pourquoi un modèle d'IA a été envisagé afin de le résoudre et a été identifié comme la meilleure solution. **Description des données** ▪ Origine des données ▪ Structure et format des données ▪ Synthèses statistiques (par ex. moyenne, médiane, variance)	Élevée

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				▪ Nombre de points de données, valeurs manquantes, etc. ▪ **Prétraitement des données** ▪ Méthodes de nettoyage des données ▪ Engineering & sélection de features ▪ Normalisation et mise à l'échelle ▪ Répartitions en données de training, de validation et de test **Choix du modèle** ▪ Description des différents modèles envisagés (par ex. réseaux neuronaux, arbres de décision, SVM) ▪ Justification du choix d'un modèle précis **Processus de training** ▪ Stratégie et paramètres de training (par ex. taux d'apprentissage, taille du lot) ▪ Algorithmes d'optimisation utilisés ▪ Méthodes de régularisation, si employées **Résultats et métriques de performance** ▪ Perte de training et de validation au fil du temps	

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				▪ D'autres métriques selon le cas d'application (par ex. précision, score F1, AUC) ▪ Surajustement (overfitting) éventuel et mesures pour y remédier **Visualisations** ▪ Courbes d'apprentissage ▪ Le cas échéant, distributions de poids et d'activation ▪ Pour les tâches de classification: matrices de confusion, courbes ROC, etc. **Discussion et interprétation** ▪ Analyse des résultats et du comportement de modèle ▪ Points faibles potentiels du modèle et propositions d'amélioration **Clôture et recommandations** ▪ Synthèse des conclusions les plus importantes ▪ Recommandations pour la prochaine itération ou mise en œuvre **Annexes** Segment de code ou référence au code complet.	

Tableau 21: Exigence systèmes d'IA – Transparence

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A18	-	Lignes directrices intelligence artificielle Confédération (transparence)	Équipe de base GS IA

Tableau 22: Exigence systèmes d'IA – Transparence informations complémentaires

8.9 Explicabilité

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
A19	Conception & développement	FACULTATIF	Lorsqu'un système d'IA présente des risques pour la personnalité et les droits fondamentaux des utilisateurs finaux ou des personnes concernées, les systèmes d'IA devraient être dotés de fonctions d'explicabilité intégrées propres à expliciter toute la logique et les motivations d'une décision de manière claire et compréhensible, tant pour les experts que pour les non-spécialistes, avec l'appui d'un training approprié. Sont notamment concernées des formations appropriées pour les Stakeholders, adaptées à leur rôle respectif.	**A19M1: Mise en œuvre de méthodes Explainable AI (XAI):** Intégration de méthodes et d'outils d'explication dans les algorithmes pouvant résumer la prise de décision dans un langage simple. Ces outils pourraient, par exemple, mettre en exergue les principales caractéristiques ayant conduit à une décision ou proposer un suivi pas-à-pas du parcours décisionnel de l'algorithme. **A19M2: Visualisation conviviale des données et décisions** Développement de visualisations interactives, qui permettent aux utilisateurs de suivre les données et les décisions de l'algorithme basées sur ces données. Par exemple, des diagrammes, des organigrammes, voire d'autres éléments graphiques pourraient être mobilisés afin d'illustrer la logique qui sous-tend les décisions et les rendre tangibles pour les utilisateurs ne pouvant se prévaloir d'un bagage technique. **A19M3: Formations continues et trainings aux systèmes d'IA**	Moyenne

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				Formations continues sur l'IA et les systèmes d'IA adaptées aux Stakeholders et à leurs rôles. Remarque: La liste des mesures ci-dessus n'est pas exhaustive et, selon le contexte du système d'IA, on peut contrôler si et lesquelles parmi ces mesures sont nécessaires.	
A20	Conception & développement	RECOMMANDÉ	Le mode de fonctionnement des systèmes d'IA devrait être présenté et expliqué avec la plus grande transparence possible.	**A20M1: Mode de fonctionnement du système d'IA** Mise à disposition d'outils ou d'interfaces qui offrent des éclairages sur le mode de fonctionnement du système d'IA. **A20M2: Système pour les demandes des utilisateurs** Recours à des mécanismes ou systèmes afin de répondre de manière efficace et axée sur l'utilisateur aux demandes de ce dernier concernant les processus décisionnels des systèmes d'IA. Les utilisateurs peuvent, par exemple, voir les données statistiques, les processus de traitement et les parcours décisionnels des systèmes d'IA. **A20M3: Documentation et outils d'aide à la prise de décision** Préparation d'une documentation et d'outils expliquant comment le système d'IA aboutit	Moyenne

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				à ses résultats. Il peut s'agir de White Papers, de FAQ ou de tutoriels vidéo fournissant des explications faciles à comprendre sur la technologie derrière le système.	
A21	Conception & développement Vérification & validation	FACULTATIF	Le système d'IA doit être soumis à une vérification visant à déterminer s'il s'agit d'un modèle White Box ou Black Box. Cette vérification doit être effectuée en partant d'approches connues permettent une classification fiable par catégorie. Fort de cette classification, on doit ensuite appliquer des mesures à vocation d'explicabilité et de transparence. Si, pour les modèles White Box, des approches d'explication basiques suffisent, les modèles Black Box en revanche nécessitent des critères spécifiques pour l'interprétation, censés garantir la traçabilité de la prise de décision. Ces critères doivent garantir que les décisions peuvent être communiquées de façon transparente et compréhensible afin d'informer correctement l'intégralité des groupes d'utilisateurs concernés.	A21M1: Contrôle Black Box et White Box Parmi les mesures recommandées, on trouve notamment l'approche d'Arrieta et al. 2019. Cette approche prévoit des critères censés garantir l'explicabilité des systèmes d'IA et ainsi contribuer à améliorer la confiance dans ces systèmes. Ces critères sont la simulabilité, la décomposabilité et la transparence algorithmique, qui sont décrits comme suit: **Simulabilité** signifie que toutes les étapes de calcul d'un modèle d'IA doivent pouvoir être évaluées en un temps raisonnable. Ce n'est plus le cas des réseaux neuronaux par exemple, qui contiennent des millions de poids, autrement dit d'opérations de calcul. **Décomposabilité** signifie que chaque partie du modèle - entrées, paramètres et calculs - admet une explication intuitive Là encore, seuls sont concernés quelques modèles d'IA, tels que les arbres de décision, dont les décisions sont fondées sur des critères clairs.	Basse

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				«Transparence algorithmique» enfin, la traçabilité de l'algorithme d'apprentissage proprement dit doit être établie. Une telle chose est possible dans le cas d'un modèle linéaire, car il existe une limite de décision dont la traçabilité est établie (une chose appartient à l'ensemble a ou à l'ensemble b). En comparaison, la limite de décision des réseaux neuronaux est très peu linéaire et n'est plus donnée explicitement. L'évaluation visant à démontrer qu'un système d'IA est de type black box ou white box peut être effectuée au moyen de la classification décrite par la figure 4 en annexe.	
A22	Conception & développement Vérification & validation	FACULTATIF	L'explicabilité d'un système d'IA requiert l'utilisation d'un ou de plusieurs frameworks.	**A22M1: Utilisation de frameworks pour l'explicabilité** Lors du choix d'un framework pour l'explicabilité de l'IA, il ne faut pas uniquement tenir compte de la technique, mais également des exigences et du public concernés. Il est important de bien choisir l'outil idoine en fonction du contexte et des besoins spécifiques, voir tableau en annexe. Exemples de frameworks ▪ SHAP (SHapley Additive exPlanations) ▪ LIME (Local Interpretable Model-Agnostic Explanations)	Basse

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				▪ What-if Tool ▪ AIX360 (AI Explainability 360) ▪ Skater	
A23	Conception & développement Vérification & validation	FACULTATIF	Le framework d'IA explicable devrait offrir la possibilité de visualiser les résultats.	**A23M1: Développement d'un outil de visualisation intégré** Mise en œuvre d'outils directement dans le framework, qui génèrent automatiquement des visualisations pour les résultats de l'IA. Il pourrait s'agir notamment de graphiques interactifs, de Heat Maps ou d'arborescences illustrant la façon dont les entrées ont mené à des décisions. **A23M2: Interfaces avec les bibliothèques de visualisation existantes** Création d'interfaces avec des bibliothèques de visualisation de données établies comme Matplotlib, Seaborn ou D3.js. Les utilisateurs du framework peuvent ainsi recourir aux outils de visualisation performants aux fins de représenter et d'interpréter les résultats et les décisions des modèles d'IA.	Basse
A24	Conception & développement Vérification & validation	FACULTATIF	Le framework d'IA explicable devrait prendre en charge les différents types de données (données en tableau, texte, images, etc.).	**A24M1: Modularisation du framework** Développement d'un framework afin qu'il contienne des composants modulaires, qui soient chacun spécialisés dans le traitement et l'analyse d'un type de données précis. Chaque module peut alors mettre en œuvre	Basse

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				des algorithmes et des méthodes de traitement spécifiques pour le type de données concerné, afin de répondre à ses spécificités et exigences. **A24M2: Interfaces extensibles pour la saisie de données** Ces interfaces devraient être conçues de façon à pouvoir être facilement complétées par de nouveaux types de données, que ce soit par l'ajout de code pour de nouveaux formats ou par l'intégration de plug-ins ou de bibliothèques conçus spécialement pour ces données.	
A25	Conception & développement Vérification & validation	FACULTATIF	Le framework d'IA explicable devrait offrir une option d'interaction avec le modèle aux fins d'un examen plus poussé.	**A25M1: Interface de Query interactive** Mise en œuvre d'une interface conviviale de Query qui permet aux utilisateurs d'adresser des demandes au modèle et de tester différents scénarios. Cette interface pourrait, par exemple, accepter les inputs de l'utilisateur afin de générer des prédictions, tout en expliquant en quoi les différentes valeurs d'entrée influent sur les résultats. **A25M2: Outils d'analyse What-if** Mettre à disposition des outils qui prennent en charge les analyses What-if. Ceux-ci permettent aux utilisateurs de visualiser l'impact des modifications apportées aux données d'entrée sur les sorties du modèle. Les	Basse

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				utilisateurs pourraient expérimenter avec les valeurs d'entrée afin de bien saisir la robustesse du modèle face aux changements et de connaître les facteurs ayant le plus d'impact sur les prédictions.	
A26	Conception & développement Vérification & validation	FACULTATIF	Le framework explicable de l'IA devrait prendre en compte la protection des données et l'anonymat dans l'analyse des données.	**A26M1: Mise en œuvre de techniques d'anonymisation des données et de privacy enhancing technologies le cas échéant** Anonymisation des données à caractère personnel en amont de l'analyse; cela peut être réalisé par des techniques telles que le masquage, le hashing ou le processus K-moyennes afin de s'assurer que la personne concernée ne puisse être identifiée. **A26M2: Contrôle des accès et Audit Trails** Définition de contrôles des accès afin de s'assurer que seuls les utilisateurs autorisés ont accès aux données sensibles. Conservez des Audits Trails qui consignent l'ensemble des accès et des actions concernant les données dans le but de garantir la traçabilité historique des accès aux données et des modifications qui y sont apportées.	Basse
A27	Conception & développement Vérification & validation	FACULTATIF	Les méthodes d'explication des systèmes d'IA devraient être validées et bien documentées dans la littérature.	**A27M1: Création de directives Best Practice** Élaboration de directives Best Practice portant sur l'application et la documentation des	Basse

N°	Cycle de vie AI	Notation	Exigence	Mesure/s	Classification
				méthodes d'explication. Celles-ci devraient être fondées sur la recherche actuelle et comporter des instructions pratiques quant à la mise en œuvre et à la vérification des méthodes d'explication. Ces directives pourraient être diffusées sous la forme de rapports techniques, de ressources en ligne ou lors d'ateliers et de formations. **A27M2: Peer review et publication** Mise en œuvre d'un processus d'évaluation par les pairs (peer review) pour que les méthodes d'explication puissent être examinées par des experts indépendants. Les méthodes devraient ensuite être publiées, accompagnées des résultats de validation, dans des revues scientifiques ou à l'occasion de conférences afin de les rendre accessibles à une large communauté scientifique.	

Tableau 23: Exigence système d'IA – Explicabilité

Informations complémentaires:

Les paramètres de traçabilité (Traceability) pour les exigences de ce chapitre sont les suivants:

	Traceability		
N°	**Explicite**		**Implicite**
	Loi	***Etude/ rapport/norme int.***	***Expert Input***
A19	-	-	Équipe de base GS IA
A20	-	-	Équipe de base GS IA
A21	-	Publication: Approche d'Arrieta et al. 2019	-
A22	-	Publication: Approche d'Arrieta et al. 2019	-
A23	-	-	Équipe de base GS IA
A24	-	-	Équipe de base GS IA
A25	-	-	Équipe de base GS IA
A26	-	-	Équipe de base GS IA
A27	-	-	Équipe de base GS IA

Tableau 24: Exigence systèmes d'IA – Explicabilité informations complémentaires

9 Aide à la mise en œuvre & exemples d'application

Ce chapitre propose une aide à la mise en œuvre pratique de la norme eCH-0272. Le but est d'aider les organisations à appliquer les exigences. Il cite des exemples propres à chaque secteur et comporte une checklist complète pour l'autoévaluation.

Administration publique: Aide à la décision automatisée

Utilisation: Un système ML aide à la vérification préalable des déclarations d'impôts.

- **Type de système:** Système basé sur le ML
- **Profil de risque:** Élevée (décision ayant des conséquences financières)
- **Objectifs de protection** Protection des données (A7), reddition de comptes (A3), équité (A6)
- **Objectifs d'acceptation** Transparence (A24), explicabilité (A26)
- **Vérification & notation:**
  - Protection des données: Élevé/impératif
  - Explicabilité & transparence moyen/recommandé
- **Mesures:**
  - Traçabilité basée sur SHAP
  - Human-in-the-loop en cas de risque
  - Audit interne avec logging

Administration publique: Chatbot pour les renseignements des autorités

Utilisation: Un chatbot génératif répond aux questions des citoyens concernant un changement de domicile, les impôts, etc.

- **Type de système:** IA générative
- **Profil de risque:** Moyenne
- **Objectifs de protection** Protection des données (A7), autonomie (A18)
- **Objectifs d'acceptation** Transparence (A24), explicabilité (A26)
- **Vérification & notation:**
  - Protection des données: Élevé/impératif
  - Explicabilité & transparence moyen/recommandé
- **Mesures:**
  - Limitation de la zone de réponse par la gestion des prompts
  - Identification des réponses automatisées
  - Possibilité de fournir du feedback (A12)

Administration publique: Modèle prédictif pour la planification du trafic

Utilisation: Un modèle statistique déclaratif prévoit la densité du trafic.

- **Type de système:** Système statistique déclaratif
- **Profil de risque:** Faible
- **Objectifs de protection** Reddition de comptes (A3), équité (A6)
- **Objectifs d'acceptation** Transparence (A24)
- **Vérification & notation:**
  - Reddition de comptes: moyen/recommandé
  - Transparence: moyen/recommandé
- **Mesures:**
  - Documentation des hypothèses de modèle
  - Traçabilité grâce aux Dashboards visuels
  - Publication des sources de données

Santé: Diagnostic par imagerie basé sur l'IA

Utilisation: ML-Tool en appui d'examens radiologiques.

- **Type de système:** Système hybride
- **Profil de risque:** Élevée (critique du point de vue médical)
- **Objectifs de protection** Autonomie (A18), droits fondamentaux (A15)
- **Objectifs d'acceptation** Transparence (A24), explicabilité (A26)
- **Mesures:**
  - Obligation d'autorisation médicale (A18M2)
  - Formation sur un jeu de données local, conforme à la LPD
  - Interface d'explication documentée (par ex. LIME)

Éducation: IA générative pour la correction d'examens

Utilisation: Le LLM local attribue une valeur préalable aux réponses en texte libre.

- **Type de système:** IA générative
- **Profil de risque:** Moyen
- **Objectifs de protection** Équité (A6), reddition de comptes (A3)
- **Objectifs d'acceptation** Transparence (A24), explicabilité (A26)
- **Mesures:**
  - Procédure de scoring documentée
  - Divulgation des modèles de prompt pour le feedback
  - Contrôle manuel par échantillonnage

10 Checklist pour la mise en œuvre d'eCH-0272

Cette checklist aide les organisations à recenser les exigences de façon systématique. Elle repose sur les 27 exigences (A1-A27) de la norme.

N°	Question / critère	Applicable (oui/non)	Remarque
A1	Le système d'IA a-t-il fait l'objet d'une classification des risques?		
A2	La méthode d'apprentissage (ML, symbolique, etc.) est-elle documentée?		
A3	Le système peut-il faire l'objet d'un audit ou d'un suivi (reddition de comptes)?		
A4	Y a-t-il un examen des impacts sur les droits fondamentaux?		
A5	Le contexte d'utilisation de l'IA a-t-il été documenté?		
A6	Des mesures ont-elles été prises afin de garantir l'équité?		
A7	Le respect de la protection des données est-il garanti?		
A8	Existe-t-il une analyse des impacts sur la protection des données (AIPD) (en cas de risque élevé)?		
A9	Une analyse des Stakeholders a-t-elle été effectuée?		
A10	La transparence vis-à-vis des personnes et organisations concernées a-t-elle été garantie?		
A11	Les résultats sont-ils expliqués aux utilisatrices et utilisateurs?		
A12	Les personnes et organisations concernées disposent-elles d'une possibilité de donner du feedback?		
A13	Une possibilité de recours est-elle documentée?		
A14	Des traitements des erreurs ont-ils été définis?		
A15	La conformité du système avec les droits fondamentaux est-elle vérifiée?		
A16	Existe-t-il une procédure d'assurance de la qualité?		
A17	Un mécanisme de suivi est-il défini?		
A18	Le processus de décision pour une validation sous surveillance humaine (si nécessaire) est-il prévu?		
A19	La manière dont l'explicabilité est établie est-elle documentée (par ex. par XAI)?		
A20	Y a-t-il des mesures de protection contre les attaques hostiles?		
A21	Le groupe cible se voit-il fournir des explications appropriées?		
A22	La participation au développement/à l'introduction est-elle prévue?		
A23	La qualité des données pertinentes pour les objectifs d'acceptation est-elle documentée?		
A24	La transparence des décisions est-elle documentée?		
A25	Existe-t-il une procédure de communication/préparation visuelle?		
A26	L'explicabilité est-elle garantie au niveau du modèle et des résultats?		
A27	L'intelligibilité linguistique est-elle assurée (groupes cibles)?		

Tableau 25: Checklist pour la mise en œuvre d'eCH-272

Remarque: La checklist englobe toutes les exigences formulées dans la norme (A1–27). D'autres recommandations facultatives tirées des annexes ou chapitres contextuels (par ex. cartographie des normes, aspects de gouvernance) peuvent y être ajoutées selon l'organisation concernée.

Cette checklist peut servir de base à un audit interne, à une analyse de maturité ou à la préparation en vue de vérifications externes. Elle devrait être régulièrement mise à jour.

11 Adéquation à la pratique des normes en matière d'IA

La présente norme est une nouvelle norme qui n'est pas encore entrée en vigueur. C'est pourquoi les exigences qui y sont définies font l'objet de vérifications ou de tests à partir de prototypes ou d'outils visant à s'assurer de leur adéquation à la pratique. Cette vérification ne fait pas partie de la présente norme, raison pour laquelle cette disposition n'a qu'une simple valeur indicative.

12 Sécurité

Il n'y a aucune considération de sécurité particulière.

13 Exclusion de responsabilité – droits de tiers

Les normes élaborées par l'Association **eCH** et mises gratuitement à la disposition des utilisatrices et utilisateurs ainsi que les normes de tiers adoptées, ont seulement valeur de recommandations. L'Association **eCH** ne peut en aucun cas être tenue pour responsable des décisions ou mesures prises par une utilisatrice ou un utilisateur sur la base des documents qu'elle met à disposition. L'utilisatrice ou utilisateur est tenu d'étudier attentivement les documents avant de les mettre en application et au besoin de procéder aux consultations appropriées. Les normes **eCH** ne remplacent en aucun cas les consultations techniques, organisationnelles ou juridiques appropriées dans un cas concret.

Les documents, méthodes, normes, procédés ou produits référencés dans les normes **eCH** peuvent le cas échéant être protégés par des dispositions légales sur les marques, les droits d'auteur ou les brevets. L'obtention des autorisations nécessaires auprès des personnes ou organisations détentrices des droites relève de la seule responsabilité de l'utilisatrice ou de l'utilisateur.

Bien que l'Association **eCH** mette tout en œuvre pour assurer la qualité des normes qu'elle publie, elle ne peut fournir aucune assurance ou garantie quant à l'absence d'erreur, l'actualité, l'exhaustivité et l'exactitude des documents et informations mis à disposition. La teneur des normes **eCH** peut être modifiée à tout moment sans préavis.

Toute responsabilité relative à des dommages que l'utilisatrice ou l'utilisateur pourrait subir par suite de l'utilisation des normes **eCH** est exclue dans les limites des règlementations applicables.

14 Droits d'auteur

Quiconque élabore des normes **eCH** en conserve la propriété intellectuelle. Elle ou il s'engage toutefois à mettre gratuitement, et pour autant que ce soit possible, la propriété intellectuelle en question ou ses droits à une propriété intellectuelle de tiers à la disposition des groupes de spécialistes respectifs ainsi qu'à l'Association **eCH** pour une utilisation et un développement sans restriction dans le cadre des buts de l'association.

Les normes élaborées par les groupes de spécialistes peuvent, moyennant mention du détenteur/de la détentrice des droits d'auteur **eCH** respectifs, être utilisées, développées et déployées gratuitement et sans restriction.

Les normes **eCH** sont complètement documentées et libres de toute restriction relevant du droit des brevets ou de droits de licence. La documentation correspondante peut être obtenue gratuitement.

Les présentes dispositions s'appliquent exclusivement aux normes élaborées par **eCH**, non aux normes ou produits de tiers auxquels il est fait référence dans les normes **eCH**. Les normes incluront les références appropriées aux droits de tiers.

Annexe A – Références & bibliographie

Titre / document	Editeur / source	Lien complet
Lignes directrices sur l'utilisation de l'intelligence artificielle au sein de l'administration fédérale (2019)	Chancellerie fédérale	https://www.bk.admin.ch/bk/fr/home/digitale-transformation/kuenstliche-intelligenz.html
Code de bonnes pratiques de la Confédération en matière de science des données et d'espaces de données dignes de confiance (2022)	Réseau de compétences des données de la Confédération	https://www.bar.admin.ch/bar/fr/home/themen/daten/datenethik.html
Recommandation de l'OCDE sur l'IA (2019)	OCDE	https://oecd.ai/en/dashboards/policy-areas/ai-principles
ISO/IEC 22989:2022 –IA – Concepts et terminologie	ISO / IEC	https://www.iso.org/standard/74296.html
ISO/IEC 42001:2023 – Systèmes de management de l'IA	ISO / IEC	https://www.iso.org/standard/81230.html
CEN-CENELEC JTC 21 – Stratégie de normes sur l'IA en Europe	CEN-CENELEC	https://www.cencenelec.eu/news-and-events/news/2021/ws-jtc21-artificial-intelligence/
Convention du Conseil de l'Europe sur l'intelligence artificielle (2024)	Conseil de l'Europe	https://www.coe.int/en/web/artificial-intelligence/draft-convention
Erklärbare Künstliche Intelligenz – Potenziale und Grenzen (2021)	Fraunhofer IESE / BMWi	https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/KI-Inno/2021/Studie_Erklaerbare_KI.pdf
Explainable Artificial Intelligence (XAI): A Survey	Arrieta, A. B. et al. (2019) / arXiv	https://arxiv.org/abs/1911.02508
Apprentissage automatique (ML)	Réseau de compétences en IA (CNAI)	https://cnai.swiss/dienstleistungen/terminologie-2/
GENAI	IBM	https://www.ibm.com/think/topics/generative-ai
Système symbolique	DataCamp	https://www.datacamp.com/blog/what-is-symbolic-ai
Système statistique déclaratif	SmythOS	https://smythos.com/ai-agents/ai-agent-development/symbolic-ai-vs-statistical-ai/
Système hybride	Information Age	https://www.information-age.com/what-is-hybrid-ai-123512053/
Transparence, explicabilité, autonomie, équité, obligation de rendre des comptes	TERMDAT Confédération	https://www.termdat.bk.admin.ch
Objectifs de protection	LPD, droits d'auteurs, directives d'IA	https://www.fedlex.admin.ch/eli/cc/2022/491/de
Objectifs d'acceptation	Code de bonnes pratiques de la Confédération	https://www.bar.admin.ch/bar/fr/home/themen/daten/datenethik.html

Titre / document	Editeur / source	Lien complet
Modèle White Box/Black Box	TERMDAT Confédération	https://www.termdat.bk.admin.ch
SHAP (SHapley Additive exPlanations)	Cornell University	https://christophm.github.io/interpretable-ml-book/shap.html
LIME (Local Interpretable Model-Agnostic Explanations)	Cornell University	https://christophm.github.io/interpretable-ml-book/lime.html
What-if Tool	Google	https://pair-code.github.io/what-if-tool/
AIX360 (AI Explainability 360)	IBM Research	https://aix360.mybluemix.net/
Phases du cycle de vie des systèmes d'IA (selon ISO/IEC 22989)	ISO / IEC	https://www.iso.org/standard/74296.html
Explainable AI (XAI) – Definition und Methodenübersicht	Cornell University	https://christophm.github.io/interpretable-ml-book/
Skater	Github	https://github.com/datascienceinc/Skater
Traceability	Cambridge	https://plato.stanford.edu/entries/information-science/

Tableau 26: Références & bibliographie

Annexe B – Collaboration & vérification

Nom	Organisation
Anna Mätzener, contributrice	Canton de Zurich
Bernard Knetsch, autre contributeur	Swisscom
Christian Werder, autre contributeur	Abraxas
Heidi Ates, contributrice	-
Mark Strauch, contributeur	Ville de Zurich
Mevlüt Polat, contributeur	ETH
Robin Pekerman, contributeur (responsable du groupe spécialisé)	Canton de Zurich
Ursulina Kölbener; contributrice	Canton Appenzell Rh.I

L'OFCOM et l'OFS ont été consultés.

Annexe C – Abréviations et glossaire

Le glossaire figure dans le présent document au chapitre 3, sous l'intitulé Terminologie.

Annexe D – Modifications par rapport à la version précédente

Il s'agit de la première version.

Annexe E – Liste des illustrations

Figure 1: Équilibre entre innovation et conformité 4

Figure 2: Processus de la norme eCH-0272 7

Figure 3: Vue d'ensemble des normes internationales 19

Annexe F – Liste des tableaux

Tableau 1: Rôles de Stakeholders et phases de cycle de vie: Propre représentation (traduite) selon DIN SPEC 92001-3:2023-08, 2023) 18

Tableau 2: Guide d'orientation des normes internationales et des normes eCH 20

Tableau 3: Objectifs de **protection** 22

Tableau 4: Objectifs d'acceptation 23

Tableau 5: Matrice d'orientation, tableaux des exigences 24

Tableau 6: Légende Traceability 25

Tableau 7: Exigence systèmes d'IA – Généralités 30

Tableau 8: Exigence systèmes d'IA – Généralités informations complémentaires 31

Tableau 9: Exigence systèmes d'IA - Droits fondamentaux 33

Tableau 10: Exigence systèmes d'IA – Droits fondamentaux informations complémentaires 33

Tableau 11: Exigence systèmes d'IA – Protection des données 35

Tableau 12: Exigence systèmes d'IA – Protection des données informations complémentaires 36

Tableau 13: Exigence systèmes d'IA – Droits d'auteur 38

Tableau 14: Exigence systèmes d'IA – Droits d'auteur informations complémentaires 39

Tableau 15: Exigence systèmes d'IA – Autonomie 43

Tableau 16: Exigence systèmes d'IA – Autonomie informations complémentaires 44

Tableau 17: Exigence systèmes d'IA – Équité 47

Tableau 18: Exigence systèmes d'IA – Équité informations complémentaires 47

Tableau 19: Exigence systèmes d'IA – Obligation de rendre des comptes 50

Tableau 20: Exigence systèmes d'IA – Obligation de rendre des comptes informations complémentaires..... 51

Tableau 21: Exigence systèmes d'IA – Transparence..... 53

Tableau 22: Exigence systèmes d'IA – Transparence informations complémentaires..... 54

Tableau 23: Exigence système d'IA – Explicabilité..... 62

Tableau 24: Exigence systèmes d'IA – Explicabilité informations complémentaires..... 63

Tableau 25: Checklist pour la mise en œuvre d'eCH-272 66

Tableau 26: Références & bibliographie 70