Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de l'économie,
de la formation et de la recherche DEFR

**Secrétariat d'Etat à la formation,
à la recherche et à l'innovation SEFRI**
Recherche et innovation

Défis de l'intelligence artificielle

Rapport du groupe de travail interdépartemental « Intelligence artificielle » au Conseil fédéral

Membres du groupe de travail interdépartemental Intelligence artificielle :

Albert Blarer (armasuisse)
Marcel Buffat (SG-DETEC)
Christian Busch (SEFRI, direction de projet)
Daniel Egloff (SEFRI)
Jon Fanzun (DFAE)
Gregor Haefliger (SEFRI, présidence)
Philipp Langer (SEFRI)
Bertrand Loison (OFS)
Thomas Luder (SFI)
Arié Malz (SG-DFF)
Eric Scheidegger (SECO)
Thomas Schneider (OFCOM)
Michael Schöll (OFJ)
Peter Schwaar (AFD)
Michael Stämpfli (OFCOM)
Verena Weber (SG-DEFR)

Table des matières

Table des figures

Liste des tableaux

Liste des abréviations

AA	Apprentissage automatique
AFC	Administration fédérale des contributions
AFD	Administration fédérale des douanes
AFF	Administration fédérale des finances
armasuisse	Office fédéral de l’armement
BEAT	Biometrics Evaluation and Testing
ChF	Chancellerie fédérale
CE	Commission européenne
CEDH	Convention européenne des droits de l'homme
CP	Code pénal suisse
Cst.	Constitution fédérale de la Confédération suisse
DAE	Direction des affaires européennes
DaziT	programme de la modernisation et de la numérisation de l'Administration fédérale des douanes
DDIP	Direction du droit international public
DDPS	Département fédéral de la défense, de la protection de la population et des sports
DEFR	Département fédéral de l'économie, de la formation et de la recherche
DEP	Digital Europe Programme
DETEC	Département fédéral de l'environnement, des transports, de l'énergie et de la communication
DFAE	Département fédéral des affaires étrangères
DFF	Département fédéral des finances

DFI	Département fédéral de l'intérieur
EPF	École polytechnique fédérale
EPFL	École polytechnique fédérale de Lausanne
EPFZ	École polytechnique fédérale de Zurich
FNS	Fonds national suisse de la recherche scientifique
FRI	Formation, recherche et innovation
GTI	Groupe de travail interdépartemental
IA	Intelligence artificielle
IdO	Internet des objets
IDSIA	Istituto Dalle Molle di Studi sull’Intelligenza Artificiale
IEEE	Institute of Electrical and Electronics Engineers
IPI	Institut fédéral de la propriété intellectuelle
LAMal	Loi fédérale sur l'assurance-maladie
LPD	Loi fédérale sur la protection des données
LRTV	Loi fédérale sur la radio et la télévision
MINT	Mathematics, information sciences, natural sciences, and technology
NCCR	National Centre of Competence in Research
NOGA	Nomenclature générale des activités économiques
OCDE	Organisation de coopération et de développement économiques
OFAC	Office fédéral de l’aviation civile
OFAG	Office fédéral de l'agriculture
OFCOM	Office fédéral de la communication
OFEN	Office fédéral de l'énergie
OFEV	Office fédéral de l'environnement
OFJ	Office fédéral de la justice
OFPP	Office fédéral de la protection de la population
OFROU	Office fédéral des routes
OFS	Office fédéral de la statistique
OFSP	Office fédéral de la santé publique
OFT	Office fédéral des transports
OGD	Open Government Data
ONU	Organisation des Nations unies
PNR	Programmes nationaux de recherche
R&D	Recherche et développement
RGPD	Règlement général sur la protection des données de l'Union européenne
SECO	Secrétariat d'Etat à l'économie
SEFRI	Secrétariat d’État à la formation, à la recherche et à l’innovation
SEM	Secrétariat d’Etat aux migrations
SG	Secrétariat général
SMSI	Sommet mondial sur la société de l'information
SNPC	Stratégie nationale de protection de la Suisse contre les cyberrisques
SRC	Service de renseignement de la Confédération
SUPSI	Scuola universitaria professionale della Svizzera italiana
swisstopo	Office fédéral de topographie
TIC	Technologies de l'information et de la communication
TST	Transfert de savoir et de technologie
UE	Union européenne
UIT	Union internationale des télécommunications
UNESCO	Organisation des Nations unies pour l'éducation, la science et la culture
UPIC	Unité de pilotage informatique de la Confédération
USI	Università della Svizzera italiana
ZHAW	Université des sciences appliquées de Zurich

Résumé

I. Défis de l’intelligence artificielle

Les nouvelles applications rendues possibles par les méthodes de l’intelligence artificielle (IA) comptent actuellement parmi les développements les plus prometteurs de la numérisation. L’intelligence artificielle permet d’ores et déjà de nombreuses applications qui ont connu un grand succès, par exemple dans les domaines de la reconnaissance d’images, du diagnostic médical, de la traduction ou de la mobilité. En tant que technologie de base, elle pourrait bouleverser l’ensemble des secteurs économiques et possède un important potentiel d’innovation et de croissance.

Alors que les bases mathématiques de l’intelligence artificielle ont été développées il y a déjà plusieurs décennies, ce sont la disponibilité d’énormes quantités de données et les progrès fulgurants de la puissance de calcul des ordinateurs qui ont rendu possible une utilisation intéressante et rentable des données au moyen des méthodes IA. On peut s’attendre à ce que les développements technologiques se poursuivent. De nouveaux domaines d’application de l’IA se dessinent déjà. Citons par exemple les applications dans les domaines du développement des médicaments, de la surveillance en temps réel des machines et des processus de production, de la cybersécurité ou encore de la recherche médicale.

Il est important que la Suisse exploite les potentiels que recèlent les nouvelles possibilités offertes par l’intelligence artificielle. Pour ce faire, il convient de créer un cadre optimal lui permettant de s’établir et de s’imposer comme l’un des pays leaders de l’innovation en matière de recherche, de développement et d’application de l’IA. Dans le même temps, nous devons répondre aux risques liés à la mise en œuvre de l’IA et prendre les mesures nécessaires en temps utile.

Situation de départ favorable

Dans un environnement marqué par une évolution technologique fulgurante, la recherche et le développement constituent une base importante pour le maintien de la compétitivité. Le présent rapport montre que la Suisse bénéficie d’une situation de départ favorable pour la recherche et le développement dans le domaine de l’intelligence artificielle. Si l’étendue de la recherche dans le domaine de l’IA est légèrement inférieure à la moyenne en Suisse (nombre de publications par habitant et part relative des publications), les instituts de recherche suisses comptent en revanche parmi les acteurs les plus importants quant aux publications de premier plan. La Suisse est également l’un des pays les plus dynamiques en matière de développement et d’application, indicateur mesuré sur la base du nombre de dépôts de brevets et de créations d’entreprises.

Malgré ce contexte favorable, les défis restent immenses au vu du rythme de développement et du potentiel considérable des technologies reposant sur l’intelligence artificielle. La recherche, l’innovation et la formation jouent un rôle central dans la réponse qui doit être apportée à ces défis. Par conséquent, les compétences disponibles dans ces domaines doivent en suivre les évolutions et être renforcées.

Objectif du rapport

Outre la rapidité du développement des nouvelles technologies IA, les principaux **défis résultent avant tout des nouvelles applications** permises par l’intelligence artificielle. D’une part, nous devons faire en sorte qu’elles soient rendues possibles, car la recherche et le développement ne peuvent pas libérer leur potentiel si des obstacles freinent la mise en œuvre des technologies qui en résultent. D’autre part, le cadre juridique doit être conçu de manière à déjouer les répercussions indésirables que la mise en œuvre de nouvelles technologies pourrait entraîner sans pour autant freiner les avancées technologiques. Il convient donc de déterminer également si la mise en œuvre concrète de l’IA aura des conséquences nécessitant une adaptation de la réglementation.

Ce rapport analyse les conditions-cadres applicables eu égard au recours croissant à l’IA, examine les défis spécifiques dans plusieurs champs d’application pour les domaines politiques relevant de l’ensemble de l’administration fédérale et s’interroge sur les adaptations nécessaires au niveau de la Confédération.

Le présent rapport est le fruit des travaux du groupe de travail interdépartemental Intelligence artificielle (GTI IA), créé par le Département fédéral de l’économie, de la formation et de la recherche (DEFR) à l’automne 2018 sur mandat du Conseil fédéral. À cet effet, le GTI IA a mis en place plusieurs groupes de travail thématiques qui ont consulté de nombreux experts externes dans le cadre de leurs travaux.

Éléments structurels des systèmes d’intelligence artificielle

Il n’existe aucune définition universelle et acceptée par tous de l’intelligence artificielle. Afin d’examiner les éventuelles mesures que doit prendre la Confédération, le présent rapport se penche non pas sur la technologie proprement dite, mais sur les applications permises aujourd’hui ou dans un avenir proche par cette technologie et sur leurs implications.

Dans cette perspective, l’IA peut être caractérisée – au lieu d’être définie – par divers éléments structurels liés à l’utilisation des applications actuelles des systèmes d’IA. Le présent rapport identifie en particulier quatre éléments structurels principaux, qui jouent un rôle plus ou moins important selon le domaine d’application. Les systèmes d’IA sont ainsi capables :

(1) d’analyser des données sous une forme que ne permettraient pas d’autres technologies dans leur état actuel en termes de complexité et de volume, notamment avec des algorithmes identifiant de manière autonome, par apprentissage, des caractéristiques statistiques pertinentes dans les données ;

(2) de faire des prédictions servant de base essentielle à des décisions (notamment décisions automatisées) ;

(3) de reproduire des aptitudes mises en relation avec la cognition et l’intelligence humaines ;

(4) d’agir de manière largement autonome sur cette base.

Certes, on retrouve certains de ces éléments sous différentes formes dans les applications non IA. Avec l’IA, c’est cependant la combinaison de ces éléments qui rend possibles des applications totalement nouvelles (p. ex. la reconnaissance faciale ou les véhicules entièrement automatiques).

Les **algorithmes de l’apprentissage automatique** (**algorithmes d’AA**) forment la technologie centrale et universelle qui permet le succès du développement des systèmes d’IA actuels. Ces algorithmes sont des **méthodes d’IA** extrêmement puissantes, dont la fonction se limite toutefois à identifier des modèles dans les données et à faire des prédictions simples. Un **système IA** est quant à lui en mesure de résoudre des problèmes complexes, qui ne pouvaient jusqu’à présent être résolus que par l’homme. Pour ce faire, les problèmes complexes sont subdivisés en une série de tâches de prédiction simples, qui peuvent être traitées séparément par un algorithme d’AA « simple ».

À l’heure actuelle, les systèmes d’IA de ce type sont développés spécifiquement pour un contexte d’utilisation particulier. Contrairement aux algorithmes d’AA qui constituent de plus en plus une technologie universelle, les connaissances spécifiques sur les domaines d’application concrets nécessaires pour combiner les composants AA des applications complexes en une solution complète ne seront donc pas automatisées dans un avenir proche et requièrent toujours une importante intervention humaine.

S'il est admis que l'IA permet de développer des applications qui s'inspirent des capacités cognitives et perceptives de l'être humain et simulent certains aspects de l'intelligence, elle est loin d'être comparable à l'intelligence humaine en l'état actuel de la technique.

Défis spécifiques de l'intelligence artificielle

Les approches de l'intelligence artificielle sont confrontées aux problèmes connus de la statistique, qui sont toutefois exacerbés dans le contexte des méthodes actuelles. Avec certaines méthodes IA, il devient ainsi **impossible de savoir** comment une prédiction ou un résultat est produit ou pourquoi un système IA apporte telle ou telle réponse à un problème concret. De plus, les applications basées sur l'IA sont tributaires de la qualité des données et des algorithmes. De ce fait, des **erreurs systématiques** dans les données ou les algorithmes (par exemple des déséquilibres non apparents, notamment en cas de sur-représentation ou de sous-représentation d'une catégorie de la population) concernant le volume ou la complexité des données utilisées ne sont souvent pas identifiés.

Ces défis ont également été qualifiés d'essentiels par les experts issus des milieux économiques et scientifiques consultés aux fins du présent rapport. Il s'agit avant tout de défis techniques, mais qui peuvent également déboucher sur des résultats problématiques, d'un point de vue juridique ou social, dans certains domaines d'application, par exemple lorsque des groupes de personnes font l'objet d'une discrimination systématique inacceptable en raison de décisions basées sur l'IA ou si, dans des domaines sensibles, le résultat d'une analyse ne peut pas être expliqué (p. ex. en cas de recours à l'IA par la justice).

Si ces problèmes peuvent être en partie amoindris par des moyens techniques, cela ne va pas sans entraîner des inconvénients. Les systèmes d'IA actuels sont optimisés pour identifier les relations de manière autonome. Un renforcement de l'explicabilité se fait donc au détriment de la performance de ces systèmes, ce qui remet en cause leur pertinence dans certains cas (p. ex. en matière de diagnostic médical).

Les **défis sont très variables selon le domaine d'application**. Par exemple, le manque d'explicabilité ne pose en principe pas de problème pour la recommandation d'un morceau de musique. En revanche, des lacunes dans l'explicabilité d'une analyse effectuée par un système IA sur le risque de récidive d'une personne présumée coupable ou condamnée restreignent les droits fondamentaux de cette personne.

Le projet de révision de la loi sur la protection des données[1] prend en compte ces défis et prévoit d'imposer diverses obligations à la personne ou à l'institution responsable quant aux décisions automatiques basées sur l'IA. Ainsi, la personne concernée par une décision automatique doit être informée de la nature de ladite décision si celle-ci s'accompagne de conséquences juridiques pour elle ou l'affecte de manière importante. La personne concernée peut en outre exiger que la décision soit contrôlée par une personne physique ou que la logique sur laquelle repose la décision lui soit communiquée.

La capacité des systèmes d'IA à **agir de manière de plus en plus autonome** met également à l'épreuve le cadre juridique actuel. Le Conseil fédéral a examiné à plusieurs reprises la réglementation en la matière et en a conclu que les textes existants sont suffisants eu égard à l'état actuel de la technologie, notamment pour ce qui concerne la responsabilité civile et pénale et le droit international privé.

Un cadre juridique général globalement adapté

Il ressort de la présente analyse que le cadre juridique général en vigueur en Suisse est en l'état globalement adapté aux nouveaux modèles de gestion et applications dans le domaine de l'IA. L'utilisa-

[1] Projet de révision totale de la loi fédérale sur la protection des données du 15 septembre 2017, FF **2017** 6815 ss.

tion des nouvelles technologies innovantes ne se fait pas dans un vide juridique, mais doit au contraire respecter le droit existant dans son intégralité. Les principes juridiques pertinents sont formulés d’une façon technologiquement neutre et sont de ce fait applicables aux systèmes d’IA. Ainsi, le cadre juridique existant autorise et limite l’usage de l’IA. Cela s’applique tout particulièrement pour les discriminations résultant de décisions basées sur l’IA. En conséquence, une **utilisation responsable** est définie par le système de valeurs sur lequel se fondent les règles de droit et garantie par l’ordre juridique.

Une refonte fondamentale du cadre juridique ne s’impose donc pas. Étant donné la forte dynamique technologique en cours, on ne peut cependant pas exclure que ce constat puisse rapidement évoluer.

Défis dans les différents domaines politiques

Même si le cadre juridique général est en principe adapté aux applications IA en l’état actuel des choses, les nouvelles applications possibles peuvent toutefois remettre en cause la **réglementation existante dans des domaines politiques spécifiques**.

Dans le cadre de son mandat, le GTI IA a effectué un état des lieux des défis liés à l’IA et touchant la Confédération. Il a identifié 17 domaines thématiques d’actualité à examiner en priorité.

Ces domaines thématiques ont été traités sous la responsabilité de l’office compétent. Les défis liés à l’utilisation de l’IA étant très variables selon le domaine thématique, les clarifications et consultations nécessaires ont été plus ou moins importantes. Alors que certains domaines thématiques étaient déjà suffisamment traités par les offices, des groupes de projet représentatifs ont dû être mis en place pour d’autres.

Au total, sept grands groupes de travail interdépartementaux ont été créés, et de nombreux experts et parties prenantes externes du monde scientifique et économique ont été consultés. Conformément au mandat confié par le Conseil fédéral, les réflexions concernant une utilisation transparente et responsable de l’intelligence artificielle devaient être prises en considération.

Dans le cadre des clarifications, les 17 domaines thématiques suivants ont été étudiés :

1. instances internationales et intelligence artificielle
2. clarification des intérêts suisses aux activités européennes concernant l’IA (programme pour une Europe numérique)
3. changements dans le monde du travail
4. l’intelligence artificielle dans l’industrie et les services
5. l’intelligence artificielle dans la formation
6. l’intelligence artificielle dans la science et la recherche
7. l’intelligence artificielle dans la cybersécurité et la politique de sécurité
8. intelligence artificielle, médias et sphère publique
9. mobilité automatisée et intelligence artificielle
10. l’intelligence artificielle dans la santé
11. l’intelligence artificielle dans la finance
12. l’intelligence artificielle dans l’agriculture
13. énergie, climat, environnement et intelligence artificielle
14. l’intelligence artificielle dans l’administration
15. développement du cadre juridique général au regard de l’intelligence artificielle
16. l’intelligence artificielle dans la justice
17. intelligence artificielle, données et droit de la propriété intellectuelle

Les défis sont en grande partie traités

Dans les différents domaines politiques, les développements de l’intelligence artificielle constituent dans certains cas des défis de taille. Selon les analyses, les clarifications et adaptations nécessaires restent importantes dans de nombreux domaines. Néanmoins, les acteurs suisses se sont déjà largement emparés de la question et y ont réagi. C’est notamment vrai pour la formation, la recherche et le secteur économique. Dans ces domaines, de nombreuses mesures ont été engagées pour répondre à ces défis. De même, l’administration fédérale s’attelle spécifiquement à la question de l’IA.

Certaines des actions *supplémentaires* requises identifiées dans le rapport concernent la nécessité d’intensifier les clarifications dans des domaines spécifiques, dont certaines doivent être menées de manière plus urgente. Par ailleurs, un nouveau domaine d’action central réside dans l’amélioration de la coordination, de la mise en réseau et du suivi face à la rapidité des développements technologiques. Au vu de la rapidité des développements, il faudrait, dans un premier temps, se fonder sur le présent rapport pour élaborer des lignes stratégiques au niveau de la Confédération. L’intelligence artificielle ne doit pas être appréhendée comme une technologie isolée mais bien plutôt comme une composante essentielle de la numérisation progressive de l’économie et de la société. Par conséquent, il convient de prendre en considération dans la stratégie globale de la Confédération en la matière (stratégie « Suisse numérique ») les mesures spécifiques qui s’appliquent aux différents secteurs des départements et des offices.

Les thématiques examinées dans le rapport ne représentent qu’un **instantané** eu égard à la dynamique fulgurante des progrès technologiques, que ce soit s’agissant du choix des domaines thématiques ou de la nécessité prévisible d’adapter la réglementation sectorielle.

Principes politiques éprouvés

Le présent rapport confirme que le principe fondamental d’une approche législative et réglementaire technologiquement neutre a également fait ses preuves dans l’environnement technologique de l’IA, marqué par une évolution rapide, difficilement prévisible pour le législateur. Là où les applications appellent une réglementation spécifique, cette réglementation doit être rédigée de sorte à s’appliquer de la même manière à toutes les technologies.

C’est pourquoi la Confédération s’efforce de continuer de mener une politique par principe technologiquement neutre, qui ne vise à promouvoir aucune technologie particulière et évite autant que possible l’adoption de réglementations spécifiques aux différentes technologies. Une telle ouverture de l’État vis-à-vis des nouvelles technologies permet d’exploiter de manière optimale le potentiel de nouvelles idées et innovations.

Du point de vue du Conseil fédéral, cette approche constitue un facteur de succès essentiel de la Suisse. S’appuyant sur ces principes éprouvés, la Suisse entend continuer à se positionner comme un pôle attractif en matière de recherche, de développement et d’utilisation des nouvelles technologies grâce à la sécurité juridique qu’elle offre, à sa réglementation efficace et à sa bonne réputation.

II. Champs d’action

Dans les domaines thématiques étudiés, le rapport fait état d’un grand nombre de mesures, d’initiatives et de clarifications qui ont déjà été lancées par la Confédération. Un ordre de priorité recense les activités qui font l’objet d’une attention particulière dans les différents domaines politiques. La majeure partie de ces mesures entrent dans le cadre d’activités existantes, d’attributions clairement définies et de procédures établies. S’ajoute à cela que certains domaines requièrent des éclaircissements *sup-*

plémentaires qui ne peuvent pas être traités dans le cadre des activités et des compétences existantes. Les champs d'action relevant de la Confédération sont regroupés en trois catégories, brièvement présentées ci-après :

A) champs d'action généraux touchant à tous les aspects de l'IA et devant être traités dans l'ensemble de l'administration fédérale ;

B) aperçu des champs d'action sectoriels et thématiques ;

C) autres champs d'action pour lesquels il n'existe pas encore de structures et/ou de compétences et pour lesquels le Conseil fédéral a par conséquent confié des mandats spécifiques, compte tenu du présent rapport.

A) Champs d'action généraux : garantir l'échange d'informations et de connaissances

Face au développement fulgurant des technologies d'IA, l'échange d'informations et de connaissances et le dialogue avec l'ensemble des parties prenantes doivent être établis et menés à un échelon supérieur. Pour la Confédération, l'échelon supérieur aux champs d'action thématiques cités en annexe 1 et au chapitre 6 réside principalement dans l'amélioration de la coordination, de la mise en réseau et du suivi face à la rapidité des développements technologiques :

- **Suivi des développements technologiques**

Le présent rapport est un instantané. Le développement des technologies IA doit être suivi avec attention. Pour ce faire, les activités de suivi de l'Office fédéral de la statistique (OFS) existantes doivent être mises à contribution et, le cas échéant, développées. Étant donné la dimension internationale de la question, le dialogue ne peut toutefois pas être conduit uniquement au niveau national. Les forums internationaux sont des instruments importants permettant d'aborder les questions fondamentales liées au développement et à l'utilisation de l'IA. La Suisse doit en outre s'impliquer dans ce suivi au sein des organisations internationales compétentes.

- **Échange d'informations et de connaissances et coordination au niveau international**

Le développement de l'IA intervient dans un environnement mondialisé. Ce développement et la réglementation afférente ne peuvent être encadrés que de manière limitée au niveau national. De surcroît, les questions de gouvernance se posent de manière très variable selon les secteurs, et les interdépendances entre des domaines politiques autrefois distincts ne cessent de se renforcer. Cela nécessite un renforcement de la mise en réseau interdisciplinaire et un échange d'informations et de connaissances au niveau national comme international.

La « **plateforme tripartite** » créée par l'OFCOM en vue de la préparation du Sommet mondial de l'ONU sur la société de l'information (SMSI) doit être mobilisée aux fins de l'échange d'informations et de connaissances et de la coordination des thématiques liées à l'IA (politique, social, économique et autres). Ouverte à toutes les organisations et les personnes intéressées, elle est dotée d'un **comité administratif composé de représentants de l'administration fédérale**, qui peut être sollicité au besoin pour coordonner les positions de la Confédération dans les instances internationales. Elle peut servir de **réseau de compétences interdisciplinaire national sur les questions et processus de portée internationale liés à l'IA**, pouvant également opérer une mise en réseau horizontale des connaissances et des expériences et, ainsi, développer des positions cohérentes de la Suisse à l'international.

- **Dialogue et intégration dans la stratégie « Suisse numérique »**

La numérisation a déclenché un processus de transformation soulevant des questions complexes qui doivent être examinées, notamment dans le domaine de l'intelligence artificielle. C'est pourquoi il convient de mettre en réseau tous les groupes d'intérêts concernés et de favoriser la collaboration de tous les niveaux de l'administration avec des représentants de l'économie, de la société civile, du monde politique et des milieux scientifiques.

L'intelligence artificielle ne doit pas être appréhendée comme une technologie isolée. Elle est une composante essentielle de la numérisation globale de l'économie et de la société. Des questions de fond de la politique en matière d'IA qui exigent un débat public doivent donc être traitées en coordination étroite avec d'autres questions relevant de la numérisation et intégrés à la stratégie « Suisse numérique ».

B) Aperçu des champs d'action sectoriels

Les enjeux posés par l'intelligence artificielle diffèrent fortement selon le domaine d'application. Dès lors, les nombreuses tâches de la Confédération s'articulent autour des différents secteurs politiques. Le tableau 1 donne une vue d'ensemble de tous les champs d'action sectoriels. Une brève description des mesures est présentée à l'**annexe 1**. Des explications détaillées par domaine politique se trouvent au chapitre 6.

Tableau 1 : Aperçu des champs d'action sectoriels prioritaires de la Confédération

Organes internationaux et IA (OFCOM, DFAE)
▪ Échange d'informations et de connaissances et coordination des positions de la Confédération au sein d'organes internationaux ▪ Renforcement de la gouvernance globale et prise en compte de l'IA dans la stratégie de politique étrangère 2020-2023 ▪ Renforcement de la Genève internationale en tant que place forte de la gouvernance numérique
Programme « Europe numérique » (SEFRI et autres)
▪ Examen de la participation suisse aux programmes « Horizon Europe » et « Europe numérique »
Changements dans le monde du travail (SECO)
▪ Suivi des conséquences de l'IA sur le marché du travail
L'IA dans l'industrie et les services (SECO)
▪ Suivi des évolutions de l'IA dans l'industrie et les services
L'IA dans la formation (SEFRI, cantons et autres acteurs concernés)
▪ Assurer la transmission des compétences nécessaires à l'utilisation de l'IA à tous les niveaux de formation ▪ Assurer une utilisation transparente et responsable de l'IA dans la formation
Utilisation de l'IA dans le domaine des sciences et de la recherche (SEFRI)
▪ Assurer la transmission des compétences dans la recherche et l'innovation dans le cadre de la politique FRI
L'IA dans le domaine de la cybersécurité et de la politique de défense (DFAE, armasuisse, SRC, Centre de compétences pour la cybersécurité, ChF, Armée, OFPP, DEFR, DDPS)
▪ Examen des implications de l'utilisation de systèmes basés sur l'IA sur la politique étrangère ▪ Examen de la cybersécurité, de la propagande et de la conduite de la guerre dans le cadre des nouvelles formes de menace induites par l'utilisation de l'IA ▪ Intégration de l'IA dans les stratégies, les plans d'action et les processus existants afin de renforcer les défenses contre les formes de menace induites par l'utilisation de l'IA ▪ Renforcement de l'anticipation du potentiel de l'IA en cybersécurité et en cyberdéfense à travers la collaboration avec les hautes écoles et l'industrie et au moyen de la recherche et des bancs d'essai
IA, médias et relations publiques (OFCOM, ChF, DFAE)
▪ Analyse de l'influence des intermédiaires de l'information et clarification des questions de gouvernance ▪ Suivi de l'évolution de l'utilisation de l'IA dans le domaine des médias

Mobilité automatisée et IA (OFROU, OFAC, OFT, swisstopo, DDPS/OFPP, DETEC)
▪ Mise en œuvre de mesures sur l'utilisation de l'IA dans des véhicules automatisés (trafic routier, aérien et ferroviaire) ▪ Mise en œuvre de mesures et développement de l'infrastructure d'échange de données pour l'IA dans le domaine de la mobilité automatisée ▪ Élaboration des bases relatives à la protection des données dans le domaine de la mobilité automatisée ▪ Création d'un cadre réglementaire sur l'IA en mobilité automatisée (autorisations, immatriculation) et encouragement de l'acceptation sociale (clarifications relatives à la tolérance de l'IA aux pannes)
L'IA dans le secteur de la santé (OFSP, Swissmedic)
▪ Examen du positionnement et des bases légales régissant le domaine des échantillons, des données et des biobanques en recherche sur l'être humain ▪ Examen de solutions possibles en matière de développement des médicaments dans le cadre de la loi sur les produits thérapeutiques
L'IA dans le secteur de la finance (DFF)
▪ Suivi de l'évolution concernant l'utilisation de l'IA par les acteurs des marchés financiers au vu des obligations réglementaires en matière de comportement ▪ Suivi de l'évolution des risques opérationnels relevant de droit de la surveillance dans les établissements financiers ▪ Suivi de l'évolution des primes d'assurances privées sous surveillance publique en cas de recours à l'IA
L'IA dans l'agriculture (OFAG)
▪ Suivi de l'évolution de l'IA dans l'agriculture
Énergie, climat, environnement et IA (OFEN, OFEV)
▪ Suivi des évolutions de l'IA dans le domaine « énergie » ▪ Suivi des évolutions de l'IA dans le domaine « environnement et climat »
L'IA dans l'administration (UPIC, OFS, offices fédéraux traitant de grandes quantités de données : AFD, OFS, AFC, OFAG et autres)
▪ Création et mise à disposition de bases de données communes dans l'administration fédérale ▪ Clarifications en vue de la création d'un réseau de compétences IA sous l'angle des aspects techniques de l'IA dans l'administration fédérale ▪ Communication renforcée sur les thèmes touchant à l'IA au sein de l'administration fédérale avec présentation des opportunités de l'IA ▪ Examen des bases légales relatives à l'utilisation de l'IA dans l'administration fédérale
Développement du cadre juridique général au regard de l'intelligence artificielle (DFAE)
▪ Examen de l'émergence d'un droit international spécifique de l'IA et de ses répercussions pour la Suisse ▪ Suivi des développements concernant l'identification des systèmes d'IA dans l'interaction avec les consommateurs
Utilisation de l'IA dans le domaine de la justice (OFJ, DFAE)
▪ Observation des évolutions sur l'utilisation de l'IA pour faciliter la prise de décision dans l'administration et la justice
IA, données et droits de propriété intellectuelle (OFCOM, OFJ, IPI, SG-DFI/OFS)
▪ Poursuite des travaux en cours sur la politique de la Confédération en matière de données (notamment stratégie Open Government Data) ▪ Poursuite des travaux en cours sur la protection des données (notamment révision de la loi sur la protection des données, mandat d'examen du Conseil fédéral concernant la portabilité, recommandations du groupe d'experts sur le traitement et la sécurité des données) ▪ Poursuite des travaux en cours sur la propriété intellectuelle (garantie que les évolutions continueront d'être soutenues dans le domaine de l'IA)

C) Autres champs d'action

Les défis de l'intelligence artificielle sont largement connus et traités dans les différents domaines politiques. Au vu de la forte dynamique technologique, il est nécessaire, dans presque tous les domaines, de suivre les développements de très près, de mener de plus amples clarifications et, le cas échéant, de procéder à des optimisations dans le cadre des activités courantes et des procédures existantes. Outre les nombreux travaux déjà engagés (cf. annexe 1), le rapport identifie également les besoins de clarifications *supplémentaires* suivantes, qui ne peuvent pas être apportées dans le cadre des activités et compétences existantes :

- **Évolution du droit international face à l'intelligence artificielle**

À l'heure actuelle, les efforts de réglementation de l'IA au plan international sont déployés non seulement par les États, mais aussi par de nombreux autres acteurs, par exemple les grandes sociétés technologiques ou les organisations internationales. Ils s'emploient à créer des règles spécifiques à l'IA, que ce soit sous la forme de normes industrielles ou de principes éthiques. Il faut examiner de manière plus approfondie la manière dont ces règles sont définies et qualifiées, dans quelle mesure elles produisent un droit international et quelles pourraient en être les répercussions pour la Suisse.

➔ Le rapport recommande que le DFAE (DDIP), avec la participation des départements concernés, présente au Conseil fédéral d'ici à fin 2020 un rapport qui mette en évidence comment les règles internationales naissent dans le domaine de l'IA, quelle doit être la classification de ces règles et dans quelle mesure elles créent du droit international et qui, le cas échéant, propose des mesures concernant le positionnement de la Suisse.

- **Intelligence artificielle, médias et sphère publique**

Les intermédiaires ont le potentiel d'instrumentaliser les applications IA à des fins commerciales ou politiques ou d'êtres eux-mêmes instrumentalisés à ces mêmes fins. De ce fait, la formation de l'opinion et de la volonté publiques peut être influencée, y compris dans le domaine politique. La thématique doit être approfondie et la possibilité d'une approche suisse de la gouvernance clarifiée.

➔ Le rapport recommande que le DETEC (OFCOM), en collaboration avec la ChF, présente au Conseil fédéral d'ici le printemps 2021 un rapport de gouvernance qui examine des mesures spécifiques sur la question et, le cas échéant, avance des propositions de mise en œuvre.

- **L'intelligence artificielle dans l'administration fédérale**

L'identification de processus dans l'ensemble de l'administration et l'accès transversal aux données constituent des conditions indispensables pour exploiter le potentiel de l'IA dans l'administration fédérale. Le développement et l'échange de connaissances et d'expériences au niveau interdépartemental sont essentiels pour un développement économique et coordonné de solutions IA dans l'administration fédérale. Les solutions fragmentées ne sont en revanche pas efficaces. Un guichet unique ou un réseau de compétences axé sur les aspects techniques de l'application concrète de l'IA au sein de l'administration fédérale pourrait être une solution. Ce guichet ou ce réseau devrait notamment remplir un rôle consultatif.

➔ Le rapport recommande que le DFF (UPIC), en collaboration avec le DFI (OFS) et avec la participation des autres départements et de la ChF, étudie la valeur ajoutée (avec notamment une analyse des besoins), la faisabilité et les ressources nécessaires d'un guichet unique ou d'un réseau de compétences, en portant une attention particulière aux aspects techniques de l'application de l'IA dans l'administration fédérale. Cet organe doit avoir un rôle consultatif pour ce qui concerne l'application de l'IA dans l'administration fédérale. En vue d'une mise en réseau complète et d'une réflexion technologique globale, il convient en outre de tenir compte d'autres technologies de la transformation numérique (p. ex. blockchain, Internet des objets, big data, etc.) et des défis liés à leur mise en œuvre.

- **Lignes stratégiques de la politique pertinente en matière d'IA**

Face au développement fulgurant des technologies d'IA et de l'ampleur des débats concernant l'utilisation de l'intelligence artificielle, la nécessité se fait sentir, du côté de la Confédération, d'élaborer des lignes stratégiques. Ces lignes stratégiques devront découler du présent rapport. L'intelligence artificielle ne doit pas être appréhendée comme une technologie isolée, mais comme une composante essentielle de la numérisation progressive de l'économie et de la société.

➔ Étant donné les défis exposés par le GTI IA, le rapport recommande que le DEFR (SEFRI), en collaboration avec le DETEC (OFCOM), élabore des lignes stratégiques en matière d'IA en vue de les présenter au Conseil fédéral d'ici le printemps 2020.

➔ Le rapport recommande que le DETEC (OFCOM) ménage une place importante à la politique en matière d'IA dans la stratégie « Suisse numérique ». Les mesures sectorielles pertinentes des départements et des offices sont également à prendre en considération dans ce cadre.

1 Mandat/contexte

Au cours des dernières années, la forte augmentation de la puissance de calcul des ordinateurs, l’accroissement exponentiel de la disponibilité des données et le développement de nouvelles méthodes IA de l’apprentissage automatique ont conduit à une renaissance d’applications de l’intelligence artificielle. Dans ce contexte, quelques développements remarquables montrent que le recours croissant à l’intelligence artificielle recèle le potentiel de transformer en profondeur l’économie et la société. Les exemples les plus connus d’applications réussies sont la traduction automatique, les machines capables de converser avec des personnes, les véhicules autonomes ou encore les logiciels de reconnaissance d’images, dont certains surpassent désormais l’homme. Au-delà de son intérêt pour l’utilisateur final, l’intelligence artificielle gagne en importance en tant que technologie de base. Les technologies de base sont des technologies transversales qui peuvent potentiellement pénétrer toutes les branches et avoir un fort impact sur la productivité dans de nombreux secteurs économiques.

Le 5 septembre 2018, au vu du rythme de développement des nouvelles applications technologiques rendues possibles par l’intelligence artificielle, le Conseil fédéral a, dans le cadre de l’actualisation de la stratégie « Suisse numérique », défini les opportunités et les défis de l’intelligence artificielle comme l’un des thèmes centraux de sa stratégie. Sur cette base, il a chargé le DEFR (SEFRI) de créer avant fin 2018 un groupe de travail consacré à l’intelligence artificielle en collaboration avec les autres départements. Le principal mandat de ce groupe de travail était de permettre l’échange de connaissances et de vues et de coordonner les positions de la Suisse sur la question au sein des instances internationales.

Le groupe de travail a également pour mission de soumettre au Conseil fédéral un rapport sur l’intelligence artificielle d’ici à l’automne 2019. Ce rapport doit notamment présenter les mesures existantes en la matière et, en se fondant sur les connaissances scientifiques disponibles, proposer une évaluation des éventuels nouveaux champs d’action. Il doit aussi mener une réflexion sur une utilisation transparente et responsable de l’intelligence artificielle.

À travers sa décision de septembre 2018, le Conseil fédéral souligne que, face aux défis posés par l’intelligence artificielle, il est nécessaire de créer des conditions-cadre favorables afin que l’utilisation de l’IA puisse apporter une valeur ajoutée économique et sociale et contribuer à une amélioration de notre vie quotidienne et à un renforcement de notre compétitivité.

Par ailleurs, le Conseil fédéral confirme que la Suisse s’engage, au plan national et international, pour la supervision et l’évaluation des conséquences de l’IA sur notre vie privée et professionnelle. Les conditions doivent en outre être aménagées de manière à ce que les systèmes algorithmiques de décision soient transparents et vérifiables, les responsabilités réglées et les systèmes utilisés respectueux de nos valeurs et de nos lois.

Structure du rapport

Au chapitre 2, le présent rapport procède dans un premier temps à une délimitation et à une qualification du concept d’intelligence artificielle et explique l’importance de son rôle pour l’économie (l’IA en tant que technologie de base). Au chapitre 3 sont présentées les possibilités nouvelles permises par les technologies de l’intelligence artificielle. Cette vue d’ensemble illustre le potentiel considérable que renferme l’IA. Le rapport aborde également les défis spécifiques à ces technologies. La compréhension des caractéristiques de l’IA et des problèmes afférents spécifiques est essentielle pour comprendre et classifier les diverses problématiques sociales et juridiques qui sont souvent associées à l’IA. Le chapitre 4 porte sur la classification des enjeux techniques du point de vue juridique et les défis que cela soulève.

Le chapitre 5 examine quelle est la position de la Suisse au niveau international en matière de recherche, de développement et d’utilisation de l’IA.

Le chapitre 6 détaille les résultats des travaux thématiques consacrés aux défis spécifiques de l'IA dans les différents domaines politiques de la Confédération. Il développe les 17 domaines thématiques que le GTI a identifiés comme étant prioritaires. Dans un premier temps, l'importance de l'utilisation de l'IA est expliquée pour chaque domaine thématique. Dans une deuxième section, les défis spécifiques concernant l'IA sont décrits, en particulier dans le domaine de compétence de la Confédération. Sont ensuite précisées les activités existantes que la Confédération ou les acteurs externes concernés ont déjà engagées pour relever ces défis. Enfin, le rapport évalue si les défis identifiés dans les différents domaines thématiques sont suffisamment traités au moyen des activités existantes (ou de la réglementation existante). Dans la négative, le rapport identifie les actions supplémentaires requises le cas échéant pour y répondre en temps utile.

2 Qualification du concept d'intelligence artificielle (IA)

Alors que ses principales bases mathématiques ont été développées il y a déjà plusieurs décennies, divers développements technologiques parallèles ont entraîné un regain d'intérêt pour l'intelligence artificielle.

La généralisation d'Internet et l'utilisation croissante des capteurs produisent d'immenses **quantités de données**, qui sont le plus souvent générées de manière automatique et gratuite dans de nombreux domaines d'application et sont disponibles comme données d'entraînement pour les méthodes de l'IA. La quantité de données d'entraînement est déterminante pour la qualité des applications IA. Au cours des 10 à 15 dernières années, les technologies de sauvegarde des données et la **puissance de calcul** ont connu parallèlement un développement fulgurant, favorisant le traitement d'énormes volumes de données à un coût abordable. Porté par ces possibilités, le développement de nouvelles méthodes permettant aux IA d'apprendre par elles-mêmes à partir de données (apprentissage automatique) a permis une utilisation intéressante et rentable des données et des méthodes.

Cette combinaison entre, d'une part, l'accessibilité d'énormes quantités de données et, d'autre part, la capacité des systèmes d'IA à apprendre et à modéliser de manière autonome des relations fonctionnelles offre de toutes nouvelles possibilités d'application que ne permettent pas les méthodes classiques. Ces développements ont donné naissance à quelques applications remarquables de l'intelligence artificielle, qui ont été utilisées avec succès dans différents domaines au cours des dernières années, si bien que l'IA est depuis longtemps entrée dans le quotidien de nombreuses personnes.

Comme on peut s'attendre à ce que les développements technologiques se poursuivent, y compris dans le domaine de la technologie des capteurs, de nouveaux domaines d'application de l'IA se dessinent déjà, par exemple en médecine, en recherche pharmaceutique, en cybersécurité ou dans le secteur financier. Une équipe de recherche a ainsi pu démontrer récemment qu'il était possible, à l'aide d'un système IA, de décoder les signaux que le cerveau envoie aux muscles du visage et, à partir de ces signaux, de produire des phrases entièrement parlées, basées sur l'activité cérébrale d'une personne[2].

2.1 Approches de la définition de l'intelligence artificielle

Il n'existe aucune définition universelle acceptée par tous de l'intelligence artificielle[3]. Les tentatives pour définir l'IA de manière globale se réfèrent généralement à l'intelligence humaine. L'IA est par

[2] Gopala K. Anumanchipalli, Josh Chartier et Edward F. Chang. « Speech synthesis from neural decoding of spoken sentences », in *Nature*, volume 568, 2019, pages 493–498. https://techcrunch.com/2019/04/24/scientists-pull-speech-directly-from-the-brain/

[3] Pour un aperçu des définitions, voir OCDE. *Artificial Intelligence in Society*, 2019, https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm

exemple définie comme suit : *intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals*.

Ce type de définition est cependant problématique en cela qu'il faudrait d'abord définir l'intelligence (notamment humaine). Or, l'intelligence est très difficile à définir avec précision. Pour remédier en partie à ce problème, on peut parler d'intelligence dès lors que celle-ci est reconnaissable en tant que telle ou qu'elle ne peut pas être distinguée d'une intelligence humaine. On parle donc d'IA par exemple lorsque des « machines exécutent des tâches faisant appel à l'intelligence chez l'humain »[4] ou lorsque le fonctionnement d'un système est considéré comme « intelligent » par l'homme[5].

Des définitions plus nuancées circonscrivent l'intelligence à certaines fonctions cognitives spécifiques (p. ex. la perception, l'apprentissage) ou y incluent d'autres aspects tels que la capacité à résoudre des problèmes complexes[6]. Tous ces aspects requièrent néanmoins d'autres définitions non triviales qui clarifient ce qu'on entend par ces concepts présupposés.

En l'état actuel de la technique, l'IA est loin d'être comparable à l'intelligence humaine. Même si les systèmes d'IA peuvent à présent exécuter des tâches complexes, réservées auparavant à l'être humain, ils sont développés pour des domaines d'application et des problématiques spécifiques et trouvent souvent leurs limites dès qu'ils sont confrontés à des problèmes que le simple bon sens humain permettrait de résoudre facilement.

Il est incontesté que les développements les plus marquants et les plus spectaculaires de l'IA sont issus d'un sous-domaine de l'informatique, à savoir des méthodes permettant aux ordinateurs d'apprendre de manière autonome (apprentissage automatique)[7].

Le présent rapport examine si le développement de l'intelligence artificielle appelle un besoin de réglementer et d'adapter le cadre juridique dans différents domaines politiques. À cet effet, les définitions qui se réfèrent aux caractéristiques de l'intelligence humaine sont inadaptées, comme détaillé à la section 3.3. Ce qui pose notamment problème avec une telle approche, c'est qu'elle efface les frontières de l'intelligence artificielle, ce qui peut déboucher sur une évaluation erronée de la technologie. S'il est vrai que les systèmes d'IA peuvent imiter de nombreuses caractéristiques de l'intelligence humaine (p. ex. l'apprentissage ou la perception), ces aptitudes ne sont pas comparables à leurs équivalents humains.

Le rapport adopte donc une approche différente : au chapitre 3, au lieu d'une définition, il délimite l'intelligence artificielle par rapport aux développements pertinents pour la politique à l'aide d'éléments structurels qui caractérisent aujourd'hui les applications IA et qui, selon le domaine d'application, jouent un rôle plus ou moins important. Ainsi, les systèmes d'IA sont capables : (1) d'analyser des données sous une forme que ne permettraient pas d'autres technologies dans leur état actuel en termes de complexité et de volume ; (2) de formuler des prédictions servant de base essentielle à des décisions (notamment décisions automatisées) ; (3) de reproduire des aptitudes mises en relation

---

[4] J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon. « A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence », 1955, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

[5] Selon le célèbre test de Turing par exemple, il ne faut pas se demander si les machines pensent, mais s'interroger sur leur capacité à démontrer un comportement intelligent identique à celui d'un être humain ou ne pouvant pas en être distingué. Voir A. M. Turing. « Computing Machinery and Intelligence », *Mind 49*, pp. 433-460, 1950. Si les systèmes basés sur des machines n'ont pas encore complètement réussi ce test, les applications actuelles n'en sont pas loin.

[6] P. ex. *a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation*. Kaplan Andreas, Michael Haenlein. « Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence », *Business Horizons*, 62(1), 2018.

[7] Voir Matt Taddy, « The Technological Elements of Artificial Intelligence», 2018, chapitre de : Ajay K. Agrawal, Joshua Gans et Avi Goldfarb (éd.). *The Economics of Artificial Intelligence: An Agenda*, NBER, à paraître ; Ian Goodfellow, Yoshua Bengio et Aaron Courville. *Deep Learning*, MIT Press, 2016, http://www.deeplearningbook.org ; Agrawal, Gans et Goldberg. « Prediction Judgment, and Complexity », 2018 ; Yann LeCun, Yoshua Bengio et Geoffrey Hinton, « Deep Learning », *Nature*, vol. 521:436-44, mai 2015.

avec la cognition et l'intelligence humaines ; et (4) d'agir de manière largement autonome sur cette base.

Cette délimitation ne constitue toutefois qu'un instantané des développements actuels qui marquent les discussions en cours. Il n'est donc pas à exclure que ces appréciations doivent être réévaluées à la suite des futurs développements technologiques et possibilités d'application.

2.2 Systèmes d'IA et méthodes de l'apprentissage automatique

Le concept d'intelligence artificielle doit être clairement distingué des *méthodes* de l'intelligence artificielle, à savoir les algorithmes de l'**apprentissage automatique (AA)**. Alors que l'expression « intelligence artificielle » est fréquemment employée dans les débats publics, de nombreux experts ne l'utilisent pas du tout et limitent l'IA à l'apprentissage automatique.

Actuellement, les applications novatrices de l'IA reposent principalement sur ce sous-domaine de la recherche IA qui permet aux ordinateurs d'apprendre de manière autonome[8]. L'AA désigne les technologies qui permettent de développer des systèmes (techniques) établissant des prévisions/prédictions dans des situations nouvelles en apprenant à partir d'expériences passées[9]. Les méthodes les plus utilisées aujourd'hui tentent aussi généralement de prévoir ou de prédire une variable à expliquer avec une série de facteurs explicatifs. Ainsi, l'AA recouvre pour l'essentiel des méthodes statistiques/algorithmiques[10] qui présentent une parenté étroite avec les méthodes statistiques établies plus connues remplissant des fonctions similaires (p. ex. les analyses de régression). L'annexe 2 présente de manière succincte le fonctionnement des algorithmes d'AA modernes.

Au contraire des anciennes approches en matière d'IA, fondées sur des règles, les processus statistiques actuels ne tentent plus de suivre des schémas humains. Les systèmes sont plutôt exercés à apprendre de leurs erreurs et à recevoir un retour sur les résultats à partir duquel ils adaptent les estimations aussi souvent qu'il le faut, jusqu'à obtenir les corrélations les plus adéquates pour le schéma à identifier. Cependant, au-delà des processus d'apprentissage, les systèmes d'IA peuvent également comprendre d'autres éléments, par exemple des éléments algorithmiques, statistiques et de régulation.

À l'inverse des logiciels plus conventionnels, qui fonctionnent toujours selon le même schéma, les décisions prises par les systèmes d'AA s'appuient sur l'optimisation. L'apprentissage automatique réalise cette optimisation en général « offline », de manière à ce que le modèle établisse le meilleur pronostic correspondant à une séquence de données connue, l'hypothèse étant que ce pronostic pourra être généralisé après l'entraînement et appliqué à de nouvelles situations. Toutefois, pour beaucoup d'énoncés de problèmes (en planification, par exemple), les situations possibles sont trop nombreuses et le modèle ne peut donc pas aboutir à un calcul prédictif. Dans ce cas, l'optimisation est réalisée dans les conditions d'utilisation réelles, le modèle générant à chaque fois plusieurs options et choisissant au final la meilleure (par exemple proposer d'autres produits à un client qui examine un produit).

[8] Bien que la plupart des développements actuels soient portés par l'AA, l'IA, en tant que sous-domaine de l'informatique, englobe de nombreuses autres approches.

[9] *Artificial intelligence is a set of techniques for developing systems that make predictions in new situations by learning from past experience*. OCDE. *Artificial Intelligence in Society*, 2019. https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm

Cette définition explique également avec justesse les récents progrès de l'IA du point de vue de l'application (cf. Agrawal, Gans et Goldberg, *Prediction Judgment, and Complexity*, 2018).

[10] Un algorithme est un code de procédure clair permettant de résoudre un problème, composé d'un nombre fini d'opérations clairement définies. Les algorithmes peuvent ainsi être implémentés dans un programme informatique pour être exécutés.

Si les **algorithmes d’AA** constituent la technologie centrale sur laquelle repose le succès du développement des systèmes d’IA actuels, ils doivent néanmoins être envisagés comme des éléments dans un contexte plus large. Les algorithmes d’AA sont des méthodes d’IA extrêmement puissantes, dont la fonction se limite toutefois à identifier des modèles dans les données et à faire des prédictions simples. Un **système IA** est quant à lui en mesure de **résoudre des problèmes complexes**, qui ne pouvaient jusqu’à présent être résolus que par l’homme. Pour ce faire, les problèmes sont subdivisés en une série de tâches de prédiction simples, qui peuvent être traitées séparément par un algorithme d’AA « simple ».

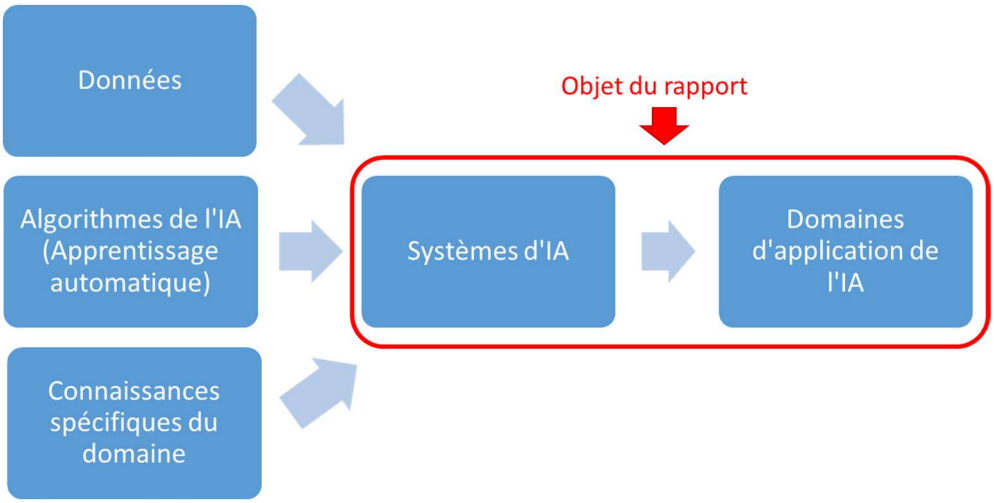
À l’heure actuelle, les systèmes d’IA plus complexes sont développés spécifiquement pour un contexte donné. Outre les méthodes d’AA, cela requiert également des connaissances spécifiques en vue de définir la structure qui subdivise un problème complexe en tâches traitables et qui organise l’interaction entre les différents composants AA. En outre, il faut régler la disponibilité et l’utilisation de données adéquates, et le système doit aussi être en mesure de s’adapter dans des conditions réelles sur la base du feed-back.

En conclusion, l’intelligence artificielle peut être **considérée comme un « système intelligent » pleinement défini**, qui, outre les méthodes employées, englobe tout un éventail d’autres éléments permettant de mettre en œuvre les méthodes d’AA de manière pertinente dans un contexte spécifique[11].

Le présent rapport traite essentiellement des conséquences de l’utilisation des systèmes d’IA dans des domaines d’application concrets (Figure 1).

[11] Matt Taddy, « The Technological Elements of Artificial Intelligence », 2018.

Figure 1 : L'apprentissage automatique, une composante des systèmes d'IA



Source : SEFRI.

Contrairement aux algorithmes d'AA qui constituent de plus en plus une technologie universelle, les connaissances nécessaires pour combiner les composants AA des applications complexes en une solution complète ne seront pas automatisées dans un avenir proche et requièrent toujours une importante intervention humaine. Plus particulièrement, des connaissances spécifiques sont nécessaires du fait que les données disponibles pour des problèmes concrets sont généralement insuffisantes pour permettre l'apprentissage d'un modèle approprié et que les taux d'erreur seraient trop élevés.

2.3 L'intelligence artificielle en tant que technologie de base

La littérature économique souligne le rôle crucial des technologies de base dans la croissance économique, la productivité et l'emploi. Les technologies de base (*general purpose technologies*) sont des technologies transversales qui ont un impact très fort sur de nombreux secteurs économiques. C'est le cas par exemple de l'électricité ou d'Internet. Les technologies de base présentent quatre caractéristiques[12,13] :

1. Elles peuvent être utilisées de manière productive dans un grand nombre de domaines d'application ;
2. Leurs prix et caractéristiques de performance évoluent fortement au fil du temps ;
3. Elles sont source d'innovations pour de nombreux produits, processus et modèles de gestion ;
4. Leur capacité à interagir avec d'autres technologies complémentaires est importante, et les évolutions qui en découlent sont nombreuses.

Comme expliqué au chapitre 3, il existe des éléments qui montrent que toutes ces caractéristiques s'appliquent aux technologies IA. Ce sont notamment les nombreuses possibilités selon lesquelles l'IA peut compléter ou remplacer les capacités cognitives et perceptives de l'être humain qui font potentiellement des technologies IA des technologies de base. Ce rôle central de l'IA est confirmé par les spécialistes scientifiques et économiques, comme décrit au chapitre 6 (sections 6.4 et 0).

[12] Cf. Commission d'experts EFI 2014 et Jovanovic/Rousseau : « General Purpose Technologies, Handbook of Economic Growth », in Aghion/Durlauf (éd.). *Handbook of Economic Growth*, Elsevier, 2005, pp. 1181-1224.

Iain M. Cockburn, Rebecca Henderson, Scott Stern. « The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis ».

Manuel Trajtenberg. *AI as the Next GPT: A Political-Economy Perspective*, 2018.

[13] Conseil fédéral. *Rapport sur les principales conditions-cadre pour l'économie numérique*, 2017, disponible à l'adresse https://www.seco.admin.ch/seco/fr/home/wirtschaftslage---wirtschaftspolitik/wirschaftspolitik/digitalisierung.html ;
SEFRI. *Défis de la numérisation pour la formation et la recherche en Suisse*, 2017, disponible à l'adresse https://www.sbfi.admin.ch/sbfi/fr/home/le-sefri/numerisation.html

Le développement et la propagation de l'IA dans l'économie, de même que ses répercussions sur l'emploi et la croissance, ne sont pas isolés du développement général de la numérisation, mais en font partie intégrante[14]. Outre les progrès dans le domaine des logiciels (ou de l'IA), les principaux moteurs de la numérisation sont les avancées en matière de puissance de calcul, de robotique ou de technologie des capteurs. La transformation numérique est également portée par les progrès dans les technologies des processus et de la mémoire ou encore le réseautage accru de l'information. Ces avancées interagissent étroitement et ne peuvent pas être dissociées. Les technologies continuent de se développer à un rythme très rapide. La probabilité de voir émerger d'autres évolutions fondamentales reste donc très élevée.

3 Caractéristiques / éléments structurels des systèmes d'IA

D'une manière générale, les technologies IA sont des outils dont l'utilisation – comme pour d'autres technologies – relève de la responsabilité de l'utilisateur. Des enjeux de société majeurs peuvent cependant résulter des applications rendues possibles par ces technologies. Du point de vue de la Confédération, l'évaluation des mesures requises doit par conséquent se concentrer sur les implications des applications possibles aujourd'hui ou dans un avenir proche des technologies IA, et non sur les technologies proprement dites.

Alors que certaines technologies et méthodes de l'IA peuvent encore, dans une certaine mesure, être délimitées de façon plausible (voir chapitre 2), la délimitation entre l'IA et d'autres technologies et applications **sur le plan de l'utilisation** reste beaucoup moins précise. Ainsi, les systèmes de réglage simples peuvent présenter une sorte de comportement intelligent (p. ex. les thermostats). D'autres applications peuvent être implémentées à partir d'autres méthodes statistiques (par exemple l'octroi de crédits ou d'assurances). De plus, la perception de ce qu'est un comportement intelligent évolue également : les machines devenant toujours plus performantes, beaucoup de tâches considérées comme intelligentes sont retirées de la définition de l'IA (« effet IA »). Par exemple, les échiquiers électroniques ne sont pour la plupart plus considérés comme faisant partie des applications de l'intelligence artificielle.

C'est la raison pour laquelle le présent rapport ne s'appuie pas en premier lieu sur une définition, mais caractérise les applications de l'intelligence artificielle au moyen de différents éléments structurels qui mettent en évidence les aspects novateurs de ces applications et les éventuelles adaptations d'ordre réglementaire (voir encadré « Vue d'ensemble des éléments structurels »). S'agissant des applications concrètes, cette approche permet également de délimiter les champs d'action politiques concernés par l'IA des autres domaines politiques (p. ex. politique des données / protection des données, réglementations sectorielles). Les éléments structurels sont détaillés ci-après[15].

Vue d'ensemble des éléments structurels

Le présent rapport identifie quatre éléments structurels principaux associés aux systèmes d'IA actuels (cf. Figure 2, champs bleus). Les systèmes d'IA sont ainsi capables :

(1) d'analyser des données sous une forme que ne permettraient pas d'autres technologies en l'état actuel en termes de complexité et de volume, notamment avec des algorithmes identifiant, par apprentissage autonome, des schémas pertinents dans les données ;

[14] Rapport en réponse aux postulats 15.3854 Reynard *Automatisation. Risques et opportunités* du 16 septembre 2015 et 17.3222 Derder *Économie numérique. Identifier les emplois de demain et la manière de stimuler leur émergence en Suisse* du 17 mars 2017, disponible à l'adresse https://www.admin.ch/gov/fr/accueil/documentation/communiques.msg-id-68708.html

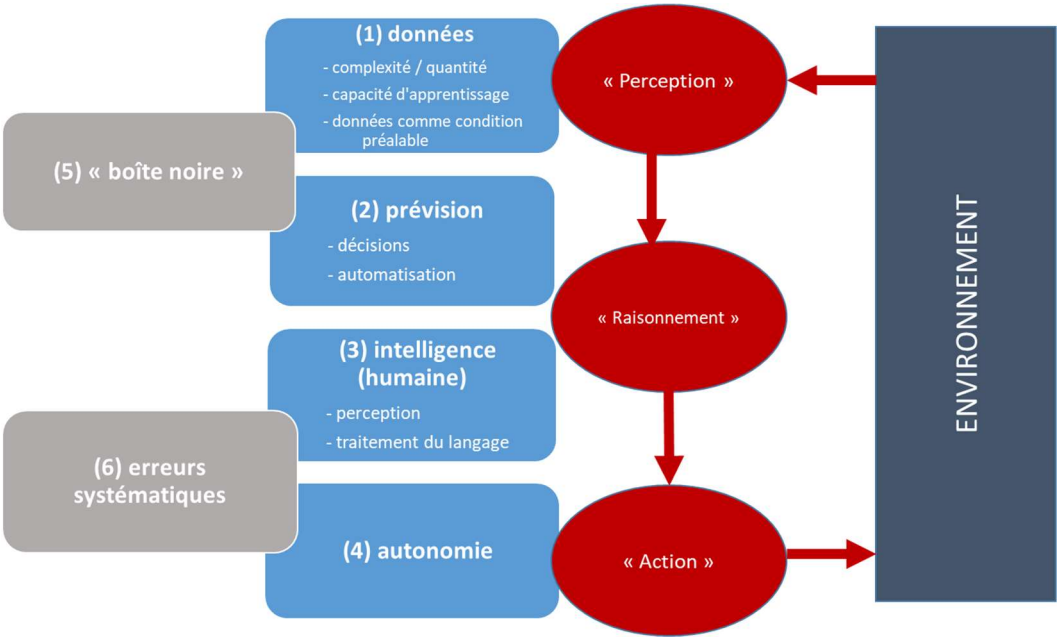
[15] Cette approche est une synthèse de plusieurs sources scientifiques. La structuration et la description ayant en outre été examinées et consolidées en concertation avec les spécialistes de l'IA proposés par swissuniversities, cette approche se fonde donc – dans la mesure du possible – sur une expertise scientifique.

(2) de faire des prédictions servant de base essentielle à des décisions (y compris des décisions automatisées) ;

(3) partant, de reproduire des aptitudes associées à la cognition et à l'intelligence humaines ;

(4) d'agir de manière largement autonome sur cette base.

Figure 2 : Éléments structurels permettant de caractériser les applications IA



Source : SEFRI.

La grande force des systèmes d'IA actuels – leur capacité à analyser d'énormes quantités de données et à apprendre de manière autonome les schémas et les relations à l'intérieur de ces données – s'accompagne également de défis techniques. Ces derniers peuvent, eux aussi, être inclus dans les éléments structurels des applications IA actuelles (champs gris de la Figure 2). Les aspects négatifs sont notamment les suivants :

(5) Il devient souvent impossible de savoir comment une prédiction ou un résultat a été produit (« phénomène de la boîte noire ») ;

(6) Il est fréquent que les relations erronées à l'intérieur des données ne puissent plus être identifiées et soient donc perpétuées (erreurs systématiques / biais).

Ces éléments structurels jouent un rôle plus ou moins important selon le domaine d'application. Tous les systèmes d'IA ne sont donc pas capables d'interagir avec leur environnement de manière autonome, loin s'en faut. Cette capacité (compte tenu des autres éléments structurels) est néanmoins nettement plus étendue dans les systèmes d'IA. La présence d'erreurs systématiques est également faible dans toutes les applications IA.

On retrouve certains de ces éléments structurels sous différentes formes dans les applications non IA (p. ex. en technique de régulation). Les caractéristiques décrites doivent en outre être analysées de manière différenciée. Pour d'autres procédés et méthodes également, l'explicabilité n'est souvent accessible qu'à un faible nombre d'experts, et pas au grand public. Mais l'IA constitue un progrès important à de nombreux égards en cela que c'est la combinaison de ces éléments structurels qui rend possibles des applications totalement nouvelles (p. ex. la reconnaissance d'images ou de langues).

Les nouvelles possibilités d'utilisation ne vont pas sans entraîner des problèmes. Cependant, ces derniers sont déjà présents sous diverses formes dans les autres méthodes statistiques (p. ex. les problèmes statistiques du biais). Mais la combinaison des différents aspects donne naissance à des défis d'une nouvelle nature (p. ex. combinaison entre le manque d'explicabilité et la possibilité de manipulation dans le cas des véhicules entièrement automatiques) ou peut accentuer fortement les problèmes (exemples : discriminations non intentionnelles en cas de difficultés à expliquer les résultats sur la base d'une distorsion des données).

Analyser la situation au vu des caractéristiques structuelles revient à réaliser un instantané. Le développement de nouveaux procédés progresse à un rythme très rapide. Le développement de méthodes permettant de mieux comprendre comment sont produits certains résultats compte parmi les domaines les plus actifs de la recherche. Il ne s'agit pas seulement de viser une plus grande transparence sociale. L'amélioration de l'explicabilité va aussi dans l'intérêt des applications et des utilisateurs.

3.1 Apprentissage à partir de données

Le terme *Big Data* fait avant tout référence aux très grandes quantités de données (dimension quantitative). Cependant, la dimension qualitative des données est au moins aussi importante : l'apprentissage automatique permet de traiter des données non structurées et qui ne sont pas disponibles avec d'autres procédés. C'est par exemple le cas des informations visuelles et vocales qui ont longtemps été considérées comme des données ne pouvant pas être traitées efficacement par ordinateur. Si les images satellites, par exemple, existent depuis plusieurs décennies, leur analyse algorithmique n'est devenue possible qu'avec les nouvelles méthodes[16].

Contrairement à d'autres méthodes statistiques, l'IA peut aussi traiter un nombre bien plus important de dimensions d'explication (*features*), permettant l'identification de relations beaucoup plus complexes qu'au moyen des méthodes statistiques classiques. Dans ce cadre, des milliers ou même des millions de dimensions d'inputs (p. ex. dans le cas du traitement d'images) peuvent être traitées de manière pertinente pour générer un résultat. Par ailleurs, il existe des applications dans lesquelles l'être humain n'est en mesure d'analyser une situation qu'avec l'aide de l'IA. Ainsi, si l'homme peut facilement s'y retrouver dans un contexte à trois dimensions, il atteint ses limites lorsque la structure des données comporte un nombre (beaucoup) plus élevé de dimensions.

Comme les méthodes d'analyses statistiques conventionnelles, qui permettent par exemple d'établir la force de la relation entre deux grandeurs (paramètre de régression) dans le cadre de l'estimation, les méthodes de l'IA déterminent elles aussi une relation adéquate entre les données d'input. L'innovation la plus décisive est la qualité de l'**apprentissage à partir de données**. Pour ce faire, soit on alimente les systèmes avec un grand nombre d'exemples déjà classés, soit on définit un ensemble de règles à l'intérieur duquel la machine apprend de manière autonome par tâtonnements.

Cela signifie que l'IA apprend elle-même à identifier des relations à partir de données brutes et d'informations. Avec les nouvelles approches, l'algorithme forme à cet effet les concepts requis de manière autonome. Ces concepts sont certes inspirés de la pensée humaine, mais ne sont que vaguement comparables à ceux que l'homme utiliserait. Par exemple, un individu caractériserait une voiture par l'existence de roues, d'un pare-brise, de rétroviseurs, de sièges, etc. Un ordinateur auquel est transmise l'image d'une voiture basée sur des pixels ne peut rien tirer de ce type de concepts abstraits. Au lieu de cela, il développe le concept de voiture, par exemple au moyen d'un grand nombre d'images de voitures, à travers des formes récurrentes d'arêtes, de lignes et de contours (cf. Annexe 2 : apprentissage automatique). La seule intervention de l'homme consiste à déterminer l'objectif pour lequel l'IA doit identifier les paramètres adéquats.

[16] Sendhil Mullainathan et Jann Spiess. « Machine Learning: An Applied Econometric Approach », *Journal of Economic Perspectives*, volume 31, n° 2, printemps 2017, pp 87-106.

Il existe une dépendance inévitable aux données, qui constitue à de nombreux égards le talon d'Achille des systèmes d'IA actuels (entraînement initial, garantie de la mise à jour des systèmes entraînés, risque de biais basé sur les données). En effet, si les systèmes d'IA sont capables de traiter efficacement de grandes quantités de données, la plupart des méthodes nécessitent inversement de grandes quantités de données pour en assurer l'entraînement. Il reste donc indispensable de disposer de données toujours plus nombreuses pour l'entraînement.

De surcroît, la **qualité des données** est déterminante. Certains algorithmes de l'apprentissage automatique, comme ceux utilisés par exemple dans la reconnaissance d'images, surpassent les capacités humaines moyennes. Pour ce faire, ils doivent néanmoins être entraînés avec de grandes quantités d'images déjà classées. Les technologies d'apprentissage automatique actuelles requièrent donc des données éditées et précises. En d'autres termes, elles ont besoin d'images dont on sait déjà qu'elles représentent un certain objet – plus généralement, un certain modèle.

Or, il est rare que les données de ce type soient facilement disponibles. La création et le recueil de données adéquates sont l'un des domaines les plus actifs de la recherche scientifique et économique sur l'IA[17] et permettent de plus en plus l'utilisation de l'IA avec des données limitées ou auparavant insuffisantes. C'est pourquoi différentes approches visent à appliquer de nouvelles formes de génération synthétique de données ou à gagner en qualité sur le traitement des données existantes. Des recherches très actives sont en outre menées dans le domaine de l'apprentissage automatique dans le but de développer des méthodes nécessitant moins de données. Beaucoup de problèmes réels sont toutefois trop complexes pour être simulés de façon réaliste et exigent toujours un accès à des données réelles.

3.2 Les prédictions à la base de décisions (automatisées)

Pour l'essentiel, le résultat d'un système IA reposant sur l'apprentissage automatique est toujours une prédiction de nature statistique – que ce soit pour la reconnaissance d'images ou vocale, la recommandation d'un produit ou la navigation[18]. Les prédictions peuvent cependant être utilisées de diverses manières. Elles constituent notamment la base essentielle de décisions.

Chaque jour, les êtres humains prennent de nombreuses décisions dans l'incertitude. Les prédictions sont une composante essentielle de ces processus de décision, car elles réduisent cette part d'incertitude. Les processus de décision peuvent donc être améliorés lorsque l'incertitude est réduite par les prédictions.

Beaucoup de problèmes peuvent être formulés sous la forme d'un problème de prédiction en vue de tirer parti des potentiels techniques de l'IA. Dans le cas d'un diagnostic médical, le médecin utilise par exemple des informations (données) sur les symptômes du patient et complète (pronostique) les informations manquantes sur la cause des symptômes. Les prédictions nécessitant le plus souvent le traitement de quantités importantes de données et la qualité d'une prédiction augmentant avec le volume de données, les méthodes de l'apprentissage automatique sont particulièrement adaptées aux systèmes conçus pour prendre des décisions sur la base de prédictions.

Si les prédictions sont une composante essentielle des décisions, ces dernières requièrent également d'autres éléments, notamment une évaluation de leurs conséquences. Les systèmes d'IA actuels sont par exemple en mesure de poser certains diagnostics de manière plus rapide et plus efficace que les

[17] OCDE. *Artificial Intelligence in Society*, 2019, p. 62.

[18] Le terme anglais *prediction* n'implique pas nécessairement une notion de temps. En statistique, il signifie qu'une variable peut être prédite au moyen d'une relation calculée par des méthodes statistiques et sur la base de valeurs données de variables explicatives. C'est pourquoi on distingue en anglais les termes *prediction* et *forecast* (prévisions pour l'avenir). En français, on emploie le terme « prédiction » pour le premier.

radiologues qualifiés (p. ex. dans le dépistage de certaines formes de cancer)[19]. Toutefois, ces systèmes ne sont que des outils et, comme les radiologues, fournissent un pronostic du type : « sur la base des informations relatives à la personne et de l'analyse des clichés d'imagerie médicale, l'augmentation du volume du foie est bénigne avec une probabilité de 66,6 %, maligne avec une probabilité de 33,3 % et inexistante avec une probabilité de 0,1% » [20].

L'interprétation et la décision quant au traitement restent néanmoins du ressort du médecin. Dans cet exemple, la décision porte sur la prescription ou non d'une intervention (p. ex. un examen invasif). Une telle décision nécessite toute une série d'évaluations supplémentaires concernant les conséquences et l'utilité de l'intervention (quelles sont les conséquences d'une intervention invasive lorsqu'une tumeur est maligne, bénigne ou inexistante ? Quelles sont les conséquences de l'absence d'intervention dans les trois cas ?).

Le recours à l'IA peut améliorer sensiblement la prédiction et présente par conséquent un vif intérêt en tant que base de décision (dans notre exemple, l'IA peut p. ex. permettre de limiter le nombre d'interventions invasives). Si les systèmes d'IA permettent d'automatiser les processus de décision, c'est toutefois dans la collaboration avec l'homme qu'ils sont le plus prometteurs, quand un individu prend une décision en se fondant sur l'input du système[21].

Réaliser une fonction d'IA de base (telle que la reconnaissance d'images) ne peut être considéré comme un choix que dans une faible mesure. Il faut au minimum une combinaison de technologies d'IA pour que des systèmes puissent prendre des « décisions » au sens étroit du terme. Dans l'exemple de la conduite automatisée, un composant d'un système d'IA peut reconnaître un panneau routier avec limitation de vitesse ; d'autres composants sont capables de déceler un véhicule lent ou de calculer, en recourant à des bases de données, la probabilité d'un tournant vers la droite. Au final, cela conduit à la décision de conduire plus lentement qu'il n'est autorisé. Mais en un sens, une pareille décision d'IA dépend de choix humains : d'abord parce qu'il a fallu faire des choix de conception en combinant les technologies d'IA pour obtenir un système d'IA ; ensuite parce qu'un être humain prend la décision d'employer un système d'IA dans un contexte donné (à la place d'un décideur humain, par exemple) ; enfin, la décision de savoir quand et comment des humains devront vérifier les résultats issus de décisions prises par des IA est aussi un choix humain[22].

3.3 Intelligence humaine

De nombreuses applications de l'IA reproduisent certains aspects des capacités cognitives et perceptives de l'être humain. Certaines d'entre elles peuvent par exemple classer des images, reconnaître des visages, traduire des textes, mais aussi effectuer des activités créatives (composer ou peindre). Quelques systèmes sont même déjà capables d'argumenter de manière convaincante dans un débat[23].

Les comportements complexes jusqu'alors réservés aux êtres humains devenant possibles avec l'IA et les technologies IA les plus avancées étant en outre vaguement inspirées de la manière dont le

[19] Cela dépend toutefois de la technique d'imagerie et du degré de standardisation des méthodes de diagnostic. Dans le cas du cancer du sein, par exemple, la standardisation est très élevée. Avec l'IRM, c'est beaucoup plus difficile, car les aimants sont légèrement différents à chaque examen, et un système dont l'entraînement est basé sur le scanner à rayons X ne peut donc pas être alimenté avec des images provenant de scanners à rayons Y.

[20] Agrawal, Ajay, Gans, Joshua et Goldfarb, Avi. « Prediction, Judgment, and Complexity: A Theory of Decision Making and Artificial Intelligence », in *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Inc., 2018.

[21] E. Strickland. « IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care », in *IEEE Spectrum*, vol. 56, n° 4, avril 2019, pp. 24-31.

[22] Cf. TA-SWISS (éd.). Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz, manuscrit non publié.

[23] CNN Business. « IBM's fast-talking AI machine just lost to a human champion in a live debate », 2019. https://edition.cnn.com/2019/02/11/tech/ai-versus-human-ibm-debate/index.html

cerveau humain apprend[24], l'IA est souvent qualifiée à tort « d'intelligence artificielle générale » humaine. Les machines seraient en mesure de prononcer des jugements, de prendre des décisions, de résoudre des problèmes variés, d'apprendre par la lecture ou l'expérience, de concevoir des concepts, de percevoir le monde et de se percevoir elles-mêmes, d'inventer et d'être créatives, de réagir à l'imprévu dans des contextes complexes et d'anticiper.

Même si les systèmes d'IA peuvent réaliser des performances cognitives surpassant l'humain et si des progrès remarquables ont été accomplis au cours des dernières années en matière de communication avec l'homme, il est admis que cette « intelligence artificielle générale » n'existe pas à l'heure actuelle. En revanche, les opinions divergent fortement quant à la question de savoir si et quand une telle IA générale pourrait voir le jour[25].

L'intelligence humaine est bien plus polyvalente et s'appuie sur un apprentissage basé sur la généralisation et l'abstraction faisant appel à différentes fonctions cognitives. Or, même si une certaine généralisation est possible, les applications de l'IA actuelles sont développées pour des domaines d'application et des problématiques spécifiques (« IA faible » ou *Narrow AI*). Les experts sollicités dans le cadre du présent rapport approuvent également cette interprétation de l'intelligence artificielle.

Mesurer l'IA à l'aune de l'intelligence humaine pose problème, car cela peut déboucher sur une évaluation erronée de la technologie. Dans l'interaction humaine, nombre de systèmes d'IA parviennent ainsi à induire en erreur l'homme avec des moyens simples, par exemple par l'intégration intentionnelle d'erreurs humaines. Dans les stratégies de réponse, ils ne s'adressent qu'en apparence à leur homologue humain. Dès les années 60 ont été développés des programmes simples qui, de façon impressionnante, ont semblé humains à des sujets volontaires pendant un court laps de temps[26]. En mai 2018, Google a présenté un système capable de passer des appels téléphoniques convaincants pour prendre des rendez-vous. Pour que le système paraisse le plus humain possible, l'IA a notamment ajouté des pauses, des approximations volontaires et des sons tels que « aha » et « hmm ».

Dans la pratique, cela peut être problématique si l'on adresse à un système des demandes qui peuvent être facilement résolues par le bon sens. On constate souvent des disparités importantes lorsque des machines font des erreurs qu'un individu n'aurait pas commises.[27] La Figure 3 montre que si l'on ajoute une perturbation minime (bruit) à l'image d'un panda – spécialement conçue pour induire en erreur le modèle de classification d'images –, l'algorithme reconnaît un gibbon avec un taux de certitude de près de 100 %, là où un individu distinguerait clairement les deux images (techniques d'apprentissage automatique contradictoire)[28].

---

[24] SATW Technology Outlook 2019. https://www.satw.ch/fr/identification-precoce/technologies/

[25] TA-SWISS (éd.). *Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz*, manuscrit non publié; OCDE. *Artificial Intelligence in Society*, 2019; Technology Review, Essay: Die sieben Todsünden der KI-Vorhersagen, 2018. https://www.heise.de/tr/artikel/Essay-Die-sieben-Todsuenden-der-KI-Vorhersagen-4003150.html

[26] Joseph Weizenbaum. ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine, 1966. http://www.cse.buffalo.edu/~rapaport/572/S02/weizenbaum.eliza.1966.pdf

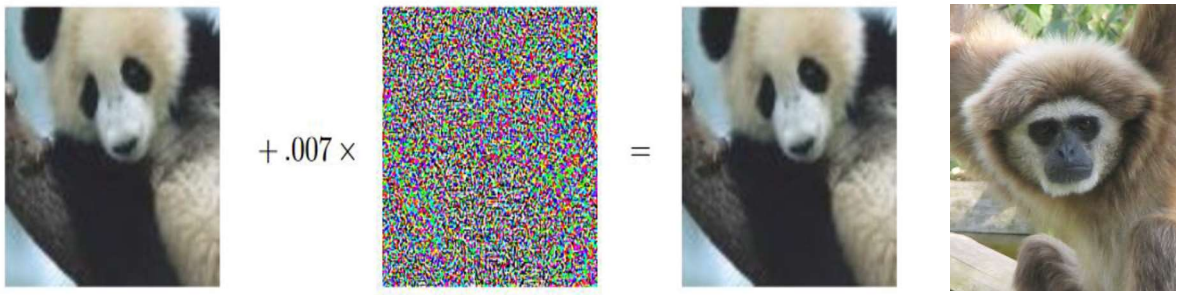
[27] Dans le « schéma de Winograd », des questions de compréhension sont par exemple posées sans qu'il n'y ait de test grammatical ou statistique simple permettant de lever les ambiguïtés :

- Les conseils municipaux ont refusé d'accorder une autorisation aux manifestants, car ils craignaient les violences.
  Qui craignait les violences ?
- Les conseils municipaux ont refusé d'accorder une autorisation aux manifestants, car ils prônaient les violences.
  Qui prônait les violences ?

Bien que les deux phrases ne se distinguent que par un mot, elles appellent une réponse opposée. Pour répondre à ces questions, il ne faut pas passer par un quelconque artifice ou la duperie, mais connaître le monde que les hommes comprennent, contrairement aux ordinateurs actuels. https://artint.info/2e/html/ArtInt2e.Ch1.S1.SS1.html

[28] OCDE. *Artificial Intelligence in Society*, p. 94, 2019.

Figure 3 : Duper les systèmes d'IA (techniques d'apprentissage automatique contradictoire)



Panda sûr à 57,7%	Nématode sûr à 8,2%	Gibbon sûr à 99,3%	À titre de comparaison : photo d'un gibbon

Source : OCDE. *Artificial Intelligence in Society*, 2019.

Des rapports ont été publiés sur ce type d'attaques dans le domaine de la reconnaissance vocale ou de la conduite automatisée[29]. Par ailleurs, des méthodes universelles permettant de produire des leurres indépendamment d'une image concrète ou d'un enregistrement sonore ont déjà été développées.

Tous les systèmes d'IA actuels sont vulnérables face aux problèmes de ce type, qui permettent de les manipuler de manière ciblée. Dans le domaine de la cybersécurité ou de la conduite automatisée (manipulation imperceptible des panneaux de signalisation), ces vulnérabilités peuvent avoir des conséquences considérables[30]. Le secteur financier peut, lui aussi, être touché, ce qui offre aux fraudeurs la possibilité de modifier le comportement des systèmes d'IA des établissements financiers en manipulant leurs données (p. ex. accorder des crédits abusifs).

Aux fins du présent rapport, une définition de l'IA inspirée de l'intelligence humaine est donc peu pertinente. Et pas seulement parce que l'état actuel de la technologie ne permet pas une telle interprétation d'une intelligence de cette sorte, mais aussi parce que l'intelligence humaine est très difficile à définir sur le plan scientifique. Enfin, l'imitation de capacités qui sont normalement attribuées à l'homme ne constitue pas non plus un phénomène nouveau (p. ex. calculatrices / échiquiers électroniques).

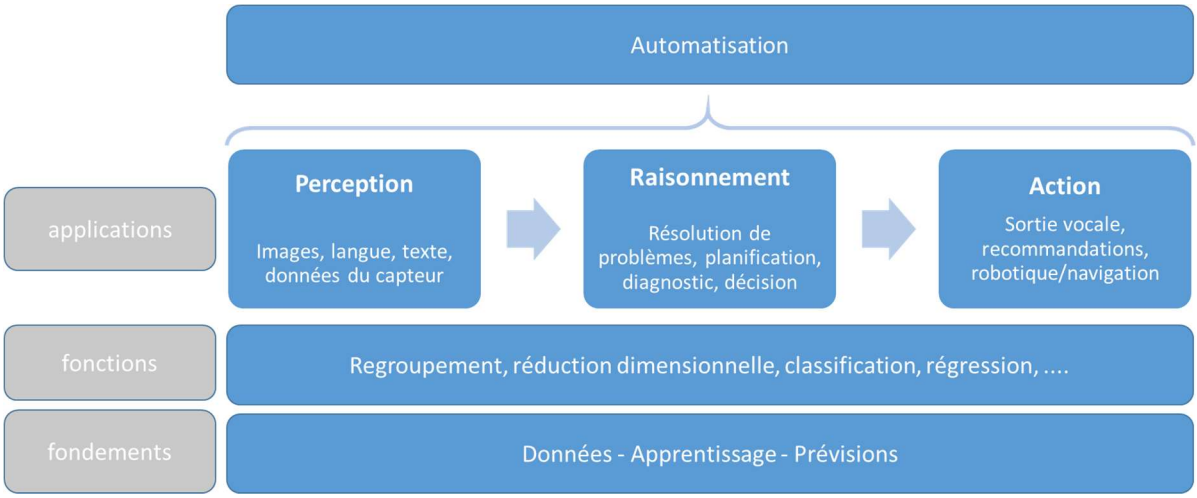
3.4 Autonomie d'action

De la capacité à traiter d'importantes quantités de données non structurées, à apprendre à partir de ces dernières et à faire des prédictions découle la capacité à « percevoir » l'environnement. Avec la capacité à transformer les prédictions en décisions, par exemple en vue de résoudre des problèmes, il en résulte que les systèmes d'IA peuvent posséder la capacité à établir une **interaction** complète – c'est-à-dire une interaction allant de la perception à l'action en passant par le traitement de l'information – et automatisée **avec l'environnement** (hommes ou machines) (Figure 4).

[29] Carlini et Wagner. *Audio Adversarial Examples: Targeted Attacks on Speech-to-Text*, 2018, arXiv:1801.01944 [cs.LG]. https://arxiv.org/abs/1801.01944 ; Yuan et al. *CommanderSong: A Systematic Approach for Practical Adversarial Voice Recognition*, 2018, arXiv:1801.08535 [cs.CR]. https://arxiv.org/abs/1801.08535 ; https://www.ics.uci.edu/~alfchen/yulong_ccs19.pdf

[30] Ackerman E. Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms, *IEEE Spectrum*, 2017. https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms ; Szegedy et al. *Intriguing properties of neural networks*, 2014, arXiv:1312.6199v4 [cs.CV]. https://arxiv.org/abs/1312.6199v4 ; Su, Vargas et Kouichi. *One pixel attack for fooling deep neural networks*, 2017, arXiv:1710.08864 [cs.LG]. https://arxiv.org/abs/1710.08864

Figure 4 : Capacités d'interaction avec l'environnement



Source : SEFRI d'après Winston P. H. Artificial Intelligence, 1992, Addison-Wesley.

Sur cette base, il existe donc une nouvelle dimension essentielle : les systèmes d'IA utilisés actuellement (ou qui seront probablement opérationnels dans un avenir proche) peuvent être employés d'une manière (ou en association avec d'autres technologies) qui permet de déduire des décisions à partir des prévisions statistiques et, à l'aide de ces dernières, d'**agir de manière autonome** dans une prochaine étape.

Le but de présent rapport est d'examiner les mesures pouvant être engagées dans les domaines intéressant la Confédération. Dans cette perspective, l'autonomie d'action est une nouvelle caractéristique essentielle des systèmes d'IA permettant de les distinguer d'autres technologies déjà très répandues. Il en découle également des implications beaucoup plus vastes pour la réglementation des différents domaines concernés. Toutefois, la prise en compte de l'autonomie d'action comme des autres critères en tant que caractéristique unique permettant de délimiter l'IA est insuffisante (dans ce cas, un grille-pain équipé d'un capteur pourrait être considéré comme un système IA).

La capacité d'autonomie peut néanmoins servir à distinguer les différentes applications IA proprement dites sur le plan réglementaire. Par exemple, les systèmes d'aide à la conduite actuels font en partie déjà appel à l'intelligence artificielle, qu'ils utilisent comme auxiliaire. Un système d'IA réellement capable de conduire de manière hautement autonome et possédant par conséquent une autonomie d'action indépendante de toute personne nécessite probablement une évaluation différente (voir sections 4.2 et 6.9).

La robotique et l'Internet des objets (IdO, ou Internet of Things) est étroitement liée à l'autonomie d'action. En robotique, les systèmes d'IA interagissent physiquement avec l'environnement ; dans les applications complexes, ce sont eux les principaux composants logiciels destinés à l'utilisation de systèmes robotiques. Une part significative de l'autonomie potentielle des systèmes d'IA repose sur le renforcement de la communication entre les machines elles-mêmes (Internet des objets). L'autonomie de l'IA ne se limite toutefois pas à l'interaction physique avec l'environnement, comme le montrent par exemple les systèmes de négoce automatisés dans le secteur financier. Ces domaines technologiques sont très proches de l'IA, mais présentent des problématiques et des enjeux qui leur sont propres[31].

La question de savoir où commence l'autonomie d'action et comment celle-ci doit être qualifiée juridiquement demeure ouverte. De même, nous devons engager un débat de société pour déterminer si, dans certains domaines, des activités identiques doivent être traitées ou non de la même manière

[31] Cf. TA-SWISS (éd.). *Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz*, manuscrit non publié.

selon qu'elles sont effectuées par un être humain ou par une machine. Des travaux empiriques conduits dans le cadre d'une étude en cours de la Fondation pour l'évaluation des choix technologiques TA-SWISS montrent qu'il existe des différences notables dans l'appréciation de la responsabilité lorsque des systèmes d'IA sont impliqués dans des processus de décision. Par exemple, on se fie beaucoup plus à un système IA conçu pour identifier les *fake news* qu'à un système IA conçu pour évaluer les candidats à un emploi. Si une personne décide qu'une décision doit être prise par un être humain ou un système IA et qu'une erreur est ensuite commise par l'être humain ou le système IA, ceux qui affichent un certain scepticisme vis-à-vis de l'IA considèreront de surcroît que la personne est beaucoup plus fautive si elle a choisi de confier la décision au système IA. À l'inverse, les spécialistes de l'IA jugeront qu'elle est beaucoup plus fautive si elle a confié la décision à un être humain.

3.5 Manque d'explicabilité

Bon nombre des méthodes IA les plus performantes aujourd'hui se caractérisent par le phénomène de la boîte noire : il devient impossible de savoir comment une prédiction ou un résultat est produit ou pourquoi un système IA apporte telle ou telle réponse à un problème concret[32]. Même si la relation entre les données d'entrée et le résultat peut être établie, une explication selon laquelle le résultat du calcul de l'IA dépend de façon non linéaire de 1000 valeurs d'entrée pondérées individuellement dans le calcul de l'IA serait certes correcte, mais d'un intérêt limité pour l'utilisateur.

D'autres défis sont associés au problème de la boîte noire. Des relations peuvent ainsi être identifiées dans les données sans qu'une théorie puisse les étayer. Un modèle peut donc fonctionner sans qu'il soit possible d'expliquer pourquoi il est efficace pour un type de problème en particulier ni comment il opère pour résoudre ce problème. Il est en outre difficile de prévoir la performance d'un modèle face au phénomène de la boîte noire[33]. S'ajoute à cela que de tels modèles ne peuvent pas être soumis à des tests complets, contrairement aux logiciels classiques. Toutes ces difficultés peuvent entraver le transfert d'un modèle du stade de développement vers la réalité. En tout état de cause, cela concerne surtout certains algorithmes d'AA de deep learning (cf. annexe 2 pour davantage d'explications). En revanche, certaines formes de l'apprentissage supervisé se rapportant à des tâches de classification débouchent sur des résultats où le processus de classification est parfaitement clair.

Le manque d'explicabilité est plus ou moins problématique selon le domaine d'application. Si l'IA est utilisée pour des traductions automatiques, le problème est peu important, puisque la qualité d'une traduction peut être assez rapidement déterminée. Dans l'exemple de la conduite automatisée, le risque qu'un système se comporte de manière totalement inexplicable dans une situation est en revanche beaucoup plus grave[34], d'autant que les erreurs de ce type sont très difficiles à anticiper du fait qu'elles sont rares. Cela impose des exigences inédites sur le plan réglementaire. Outre l'agrément technique actuellement en vigueur, un véhicule commandé par IA devrait donc être également évalué quant à la capacité du robot de conduite à assurer un fonctionnement automatique sécurisé du moyen de transport (chapitre 6.9).

De nombreux autres domaines abordés dans le présent rapport sont, eux aussi, concernés par la problématique de l'explicabilité. Par exemple, comment un médicament créé par des algorithmes doit-

[32] MIT Technology Review. *The Dark Secret at the Heart of AI*, 2017. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/
Les nouvelles technologies d'AA, en particulier les *deep neural networks*, font appel à d'autres techniques de programmation que les algorithmes informatiques « classiques ». Au lieu de structures logicielles claires, qui sont en principe transparentes au moins pour les programmeurs, ces derniers prédéfinissent certes un réseau neuronal, mais dont la connectivité et la pondération des relations évoluent sur un nombre considérable de cycles d'entraînement (p. ex. un algorithme de reconnaissance d'images est entraîné avec des millions d'images). Cf. TA-SWISS (éd.). *Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz*, manuscrit non publié.

[33] TA-SWISS (éd.). *Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz*, manuscrit non publié.

[34] MIT Technology Review. *The Dark Secret at the Heart of AI*, 2017.

il être évalué ? Comment le recours à l'IA dans le système judiciaire peut-il être concilié avec le droit d'être entendu ?

3.6 Erreurs systématiques (biais) et causalités apparentes

Le manque d'explicabilité des résultats de l'IA peut accentuer d'autres problèmes connus de la statistique : les causalités apparentes et les erreurs systématiques (« biais »).

Les **erreurs systématiques** surviennent quand les valeurs estimées s'éloignent systématiquement des « vraies » valeurs. Cela peut se produire si par exemple l'échantillon provient d'un groupe de population non représentatif. Mais une application incorrecte ou des restrictions dans la méthode d'estimation employée peuvent aussi générer des erreurs systématiques. Contrairement à ce qui se passe en cas d'erreurs fortuites, l'effet des erreurs systématiques sera d'autant plus important que le nombre d'unités analysées est élevé (notamment si l'on augmente la taille de l'échantillon).

On parle de **causalité apparente** lorsqu'il existe une relation mesurée statistiquement entre deux grandeurs (corrélation) sans qu'une relation causale soit établie. Des causalités apparentes peuvent s'observer lorsque deux grandeurs dépendent d'une cause commune qui n'est pas prise en compte[35] ou lorsqu'une corrélation est purement fortuite (coïncidence)[36].

Étant donné que les causalités apparentes existent dans les importantes quantités de données du seul fait des règles du hasard et que l'IA peut présenter le potentiel le plus élevé en la matière avec les immenses volumes de données en jeu, on peut s'attendre à ce que le problème se renforce dans le cadre des applications IA. Et ce notamment lorsque l'IA doit identifier des modèles dans d'énormes quantités de données. Mais contrairement à l'homme, une application IA n'est en principe pas en mesure de distinguer une causalité apparente d'une causalité réelle en se fondant sur des réflexions théoriques et intuitives et donc d'identifier des problèmes éventuels[37].

Il apparaît qu'une grande partie (environ la moitié selon une étude) des systèmes d'IA ont recours à des stratégies de résolution qui, du point de vue de l'homme, sont naïves[38]. Ainsi, certains systèmes d'IA analysent les images à l'aide du contexte et classent, par exemple, des images dans la catégorie « train » dès lors que des rails y apparaissent. Même si la majorité des images sont correctement classées, ces systèmes n'accomplissent pas la tâche proprement dite consistant à identifier des trains. D'autres systèmes d'IA fondent leur décision de classification sur des artefacts apparus lors de la préparation des images et n'ayant rien à voir avec le contenu des images à analyser.

Il faudrait donc pouvoir éviter que soient apprises des propriétés des modèles qui sont certes corrélées d'une manière ou d'une autre avec le résultat dans les données d'entraînement, mais ne peuvent pas être utilisées pour la prise de décision dans d'autres situations[39]. Le recours à ce type de

[35] Exemple : les compétences en calcul des enfants sont corrélées avec la longueur de leurs bras. Cette corrélation se fonde sur le fait que les enfants plus âgés à la fois savent mieux calculer et ont des bras plus longs.

[36] En Allemagne, le nombre d'examens annuels d'avocat est corrélé avec la surface forestière. Les deux grandeurs augmentent, mais le fait que leur progression suive approximativement le même rythme est uniquement le fruit du hasard.

[37] Plusieurs approches de l'AA visent à établir les relations causales. Mais elles en sont encore au stade de la recherche à l'heure actuelle.

[38] Sebastian Lapuschkin, Stephan Wäldchen, Alexander Binder, Grégoire Montavon, Wojciech Samek et Klaus-Robert Müller. « Unmasking Clever Hans predictors and assessing what machines really learn », *Nature Communications*, 2019, volume 10, article number: 1096 (2019).

[39] Le dilemme biais/variance est un problème crucial de l'apprentissage supervisé. Idéalement, il faudrait pouvoir choisir un modèle qui à la fois identifie avec précision les lois régissant les données d'entraînement et peut être généralisé à des nouvelles données de test. Cependant, il est généralement impossible d'y parvenir simultanément (cf. Scott Fortmann-Roe). *Understanding the Bias-Variance Tradeoff*, 2012. http://scott.fortmann-roe.com/docs/BiasVariance.html

systèmes mal entraînés dans le diagnostic médical ou dans des domaines critiques pour la sécurité pourrait comporter des risques considérables.

Dans les applications quotidiennes, le problème du biais peut en outre se manifester à travers des déséquilibres ou des préjugés non apparents influençant le comportement de l'algorithme. Ces aspects sont souvent abordés dans le contexte des réflexions sur **l'équité et la discrimination** des applications IA. Par exemple, il est apparu qu'une recherche Google *professional hair* donne majoritairement comme résultats des coiffures de femmes blanches, alors que la recherche *unprofessional hair* débouche sur une majorité de coiffures de femmes noires[40]. Autre exemple connu : la société Amazon a renoncé à utiliser un système algorithmique de sélection de candidats après que l'on s'est rendu compte que ce système avorisait les candidats de sexe masculin sur la base de données d'apprentissage biaisées.

Il est important ici de faire la distinction entre les aspects techniques et les aspects normatifs. En effet, un biais d'ordre statistique n'est pas équivalent à un biais au sens d'un préjugé ou d'un parti pris pour ou contre une chose, une personne ou un groupe, par rapport à d'autres choses, personnes ou groupes, d'une manière qui est en général perçue comme injuste.

Dans la pratique, le biais historique est généralement déterminant. Le plus souvent, celui-ci ne doit toutefois pas être envisagé comme une erreur d'un système IA à proprement parler, mais davantage comme un défi juridique. Si la solvabilité de certaines catégories de la population est historiquement mauvaise, les systèmes ne voient pas d'erreur, du point de vue technique, à poursuivre dans cette pratique historique[41].

Le problème du biais est particulièrement aigu dans le domaine de l'IA, car ni les programmeurs ni les utilisateurs ne sont en mesure d'identifier les déséquilibres non apparents dans les jeux de données d'entraînement lorsque ceux-ci sont composés de millions de points de données. Si l'usage de ces applications de l'IA se généralise, cela présente le risque que des personnes subissent des discriminations inadmissibles et systématiques sur la base de décisions d'IA.

Il est bien sûr possible de prendre des dispositions concernant les aspects liés à l'équité et à la discrimination. Les algorithmes peuvent ainsi être conçus de sorte qu'ils ignorent systématiquement certaines caractéristiques des données (p. ex. les informations relatives au sexe ou au statut social). Cependant, la précision des algorithmes en est affectée[42] . Le raisonnement mathématique montre qu'il n'est pas possible de satisfaire simultanément certaines exigences, telles que l'équité et la précision, avec ce type d'algorithmes[43]. De plus, la possibilité existe que des caractéristiques non souhaitées (ou non intentionnelles) puissent être déduites d'autres caractéristiques auxquelles elles sont corrélées.

La problématique de l'équité des algorithmes est complexe car les normes juridiques applicables doivent être traduites dans un « langage » compréhensible pour les programmes informatiques. Le problème est que l'idée de proposer « le même traitement aux membres de différents groupes » se prête à plusieurs interprétations possibles. Ainsi, les systèmes d'IA peuvent être programmés de manière à attribuer aux membres de différents groupes (par exemple les hommes et les femmes) la même probabilité de prévision positive (par exemple bonne candidature à un poste).

---

[40] The Guardian. « Women must act now, or male-designed robots will take over our lives », 2018. https://www.theguardian.com/commentisfree/2018/mar/13/women-robots-ai-male-artificial-intelligence-automation;

[41] Haas Berkeley. « Minority homebuyers face widespread statistical lending discrimination, study finds », 2018. https://newsroom.haas.berkeley.edu/minority-homebuyers-face-widespread-statistical-lending-discrimination-study-finds/

[42] TA-SWISS (éd.). *Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz*, manuscrit non publié.

[43] Cela est lié au problème qui veut que, dans une classification, la minimisation d'une erreur découlant d'une affectation erronée (p. ex. e-mail classé comme spam alors qu'il ne s'agit pas d'un spam) se fait au détriment de celle d'une non-affectation erronée (l'e-mail n'est pas classé comme spam alors qu'il s'agit d'un spam).

Toutefois, on pourrait tout aussi bien exiger du système qu'il ne fasse pas de distinction entre les erreurs de classification de premier et de deuxième rang (x classé par erreur – ou au contraire non classé – dans le groupe G) sur des groupes sensibles à la discrimination. On pourrait par exemple exiger qu'un système d'IA sur l'évaluation de candidatures accorde la même probabilité au fait de retirer par erreur des femmes qualifiées que des hommes qualifiés – de telles erreurs de classification se produiront toujours.

4 Aspects généraux et qualification juridique

Selon l'application, les enjeux techniques liés à l'utilisation de l'IA décrits au chapitre précédent ont des implications juridiques concrètes. Toute évaluation réalisée par la Confédération doit donc impérativement s'interroger sur la qualification juridique de ces problématiques. La réponse ne saurait être générale, mais doit au contraire s'intéresser aux cas d'application concrets. Par exemple, les exigences en termes de transparence et d'explicabilité qui s'appliquent à un système IA faisant des recommandations de restaurants ne sont pas les mêmes que celles applicables à un système IA utilisé par la justice ou se déplaçant dans le trafic sans conducteur humain.

On peut cependant faire quelques observations générales sur la manière dont les défis afférents aux applications IA affectent notre système juridique et de valeurs actuel. Aussi la section 4.1 s'attachera-t-elle dans un premier temps à démontrer que **le cadre juridique en vigueur et les principes fondamentaux** de la Confédération en matière de nouvelles technologies s'appliquent également à l'utilisation de l'intelligence artificielle. Les sections 4.2 à 4.4 examineront ensuite les enjeux techniques spécifiques des méthodes de l'intelligence artificielle (accroissement de l'autonomie, manque d'explicabilité, erreurs systématiques) et décriront comment ceux-ci doivent être qualifiés **sur le plan juridique** (cf. Tableau 2).

Tableau 2 : Principaux enjeux techniques et juridiques de l'IA

Enjeu technique		Enjeu juridique
Autonomie d'action	→	Responsabilité
Phénomène de la boîte noire	→	Transparence/explicabilité
Biais / erreur systématique	→	Discrimination possible
Disponibilité et qualité des données	→	Accès aux données et protection des données

Source : SEFRI.

4.1 Principes de la politique de la Confédération en matière de nouvelles technologies

Face au rythme fulgurant des avancées technologiques, le Conseil fédéral s'est à plusieurs reprises exprimé sur les principes de la politique de la Confédération s'appliquant aux technologies numériques nouvelles et innovantes[44]. Le présent rapport se fonde également sur ces principes :

(i) Approche *bottom-up*
La politique doit veiller à l'instauration de conditions-cadre optimales et propices à l'innovation permettant l'essor des nouvelles technologies, tandis que la décision quant au choix des technologies spécifiques doit être laissée aux acteurs concernés. Même dans le cadre de

[44] Rapport du Conseil fédéral. *Bases juridiques pour la distributed ledger technology et la blockchain en Suisse* ; numérisation ; *Une politique industrielle pour la Suisse*, rapport rédigé le 16.04.2014 en réponse au postulat Bischof.

l'encouragement de la recherche et de l'innovation, la Confédération ne souhaite pas promouvoir de technologies spécifiques.

(ii) Point de vue de l'utilisation

L'évaluation des nouvelles technologies porte prioritairement sur les applications et leurs répercussions. L'examen des réglementations ne doit pas être axé sur la technologie proprement dite, mais se concentrer sur les lacunes existant dans le domaine des **applications concrètes** de l'intelligence artificielle et sur les risques que celles-ci posent pour les droits fondamentaux des personnes concernées. Il convient en premier lieu d'éliminer les obstacles entravant l'utilisation des nouvelles technologies. Dans le même temps, les conséquences négatives de cette utilisation doivent être atténuées ou empêchées.

(iii) Neutralité technologique

En matière de législation et de réglementation, la Suisse suit une approche fondée sur des principes et neutre sur le plan technologique, tout en permettant les exceptions nécessaires ; ce faisant, les règles créées doivent être aussi neutres que possible du point de vue de la concurrence. Les dispositions légales ne doivent pas viser certaines technologies en particulier. Au contraire, elles doivent en principe (c'est-à-dire lorsque cela est possible et judicieux) traiter de la même manière des activités et des risques comparables.

(iv) Défaillances du marché

Pour des raisons d'efficacité, l'État ne doit réglementer le marché que si son intervention permet d'en accroître l'efficience par rapport à la situation actuelle. En principe, le marché règle de manière optimale le problème de la coordination des différentes activités économiques et des intérêts divergents des acteurs. L'État ne devrait donc prendre des mesures que pour remédier aux défaillances du marché. En l'absence de défaillance du marché – ou d'autres intérêts publics prépondérants – et si l'utilisation de l'IA intervient dans le cadre d'activités économiques privées, aucune réglementation ne doit en principe être promulguée.

(v) Admissibilité juridique

Il découle de l'approche *bottom-up* que l'utilisation de l'IA – tout comme le choix d'autres technologies – ne justifie en soi aucune mesure de l'État ni réglementation.
Comme toute autre technologie, les systèmes d'intelligence artificielle sont avant tout des outils dont l'utilisation par les particuliers est en principe autorisée. L'utilisation de technologies innovantes n'intervient toutefois pas dans un vide juridique, mais doit se conformer à l'ensemble du droit en vigueur. La question de la réglementation se pose notamment lorsque les applications d'IA touchent aux droits fondamentaux, risquent d'entraîner une défaillance du marché ou concernent l'action de l'État.

(vi) Attention particulière portée aux droits fondamentaux et aux droits de l'homme

Les droits fondamentaux sont des droits élémentaires de l'individu et constituent des valeurs et principes d'organisation centraux de l'État de droit. C'est pourquoi les buts visés par les droits fondamentaux doivent être respectés dans l'ensemble de l'ordre juridique, et il faut veiller à leur pleine réalisation. Outre l'exercice de droits subjectifs vis-à-vis de l'État, cet objectif est aussi notamment atteint au moyen de la dimension programmatique des droits fondamentaux. Il incombe ainsi aux organes législatifs de promulguer des textes qui garantissent au mieux la liberté et l'égalité de traitement en prenant en compte les droits fondamentaux[45]. L'État doit en outre veiller à ce que les droits fondamentaux, dans la mesure où ils s'y prêtent, soient aussi réalisés dans les relations qui lient les particuliers entre eux (art. 35, al. 3, Cst.). Cela inclut également les droits de l'homme garantis par le droit international, telle la Convention européenne des droits de l'homme du 4 novembre 1950 (CEDH ; RS 0.101).

L'IA peut avoir des répercussions sur la plupart des droits fondamentaux et des droits de l'homme. Si ces droits en sont affectés ou si l'ordre juridique en vigueur s'avère insuffisant, une

[45] Regina Kiener, Walter Kälin et Judith Wyttenbach. *Grundrechte*, 3e édition, 2018.

réglementation est nécessaire. Dans ce contexte se pose la question de savoir s'il faut promulguer des prescriptions réglementaires (minimales), prendre des dispositions institutionnelles et instaurer un mécanisme de contrôle de l'État de droit ou si l'on peut s'appuyer sur des directives éthiques ou techniques. D'une manière générale, il convient de s'assurer que l'IA se développe pour le bien de la société et est utilisée en ce sens.

(vii) Base légale requise pour l'action de l'État

En principe, l'État (administration, justice) est autorisé à utiliser l'IA comme un outil, même lorsque le statut juridique de personnes en est affecté, pour autant qu'il existe une base légale qui lui permette d'agir selon des modalités concrètes ou techniques. Il existe une sensibilité particulière dans ce domaine pour ce qui est du respect des droits fondamentaux et des droits de l'homme, car ces derniers concernent en premier lieu les organes étatiques.

De l'avis du Conseil fédéral, il n'appartient en principe pas aux autorités de décider quelle technologie va s'imposer et dans quelle mesure. Cette approche a déjà fait ses preuves, surtout dans un environnement technologique en mutation rapide et dont l'évolution n'est que partiellement prévisible pour le législateur. Premièrement, elle offre une grande flexibilité. Deuxièmement, elle est conforme à l'objectif de neutralité en matière de concurrence. Troisièmement, une approche neutre sur le plan technologique permet de résoudre un problème potentiel, à savoir que les processus législatifs durables sont souvent à la traîne du progrès technologique. Il ne faut toutefois pas exclure de prévoir des exceptions dans certains domaines dans lesquels une adaptation juridique spécifique des technologies d'IA serait indiquée.

Sur la base de ces principes, des conditions-cadre optimales doivent être créées afin que la Suisse puisse se positionner comme un pôle attractif en matière d'utilisation des nouvelles technologies. C'est pourquoi le processus de transformation numérique qui se profile doit obéir à une approche transversale, en dialogue avec toutes les parties prenantes, afin de permettre à tout un chacun de saisir les opportunités liées aux avancées numériques et d'en surmonter les difficultés. Le Conseil fédéral a inscrit ces principes dans sa stratégie « Suisse numérique ». Une telle ouverture de l'État vis-à-vis de toutes les technologies, et surtout des nouvelles technologies, permet d'exploiter le potentiel de nouvelles idées et innovations.

4.2 Autonomie et responsabilité

La capacité des systèmes d'IA à agir de manière de plus en plus autonome met à l'épreuve le cadre juridique actuel. À l'heure actuelle, la question de la qualification juridique de l'autonomie concerne principalement l'utilisation de l'IA dans le domaine de la robotique. Du point de vue du droit civil se pose la question de savoir qui doit assumer la responsabilité des dommages causés par un système de ce type. Le droit suisse de la responsabilité civile ayant un caractère très général, il est souple et technologiquement neutre. Les règles générales de responsabilité peuvent également être appliquées aux nouvelles technologies, les robots étant en principe considérés comme des choses dans ce contexte. Dans tous les cas, seule une personne physique ou morale – et non la machine – peut voir sa responsabilité engagée. La responsabilité de l'exploitation de systèmes informatiques autonomes doit toujours relever des actes ou des omissions d'une personne responsable, y compris lorsque la machine agit sans la supervision directe de ladite personne[46].

En outre, en l'état actuel des choses, il est difficile d'attribuer à des systèmes d'IA agissant de façon autonome les caractéristiques nécessaires à la détermination de la responsabilité dans les rapports juridiques. Les machines ne semblent ni en mesure d'agir intentionnellement (c.-à-d. avec conscience et volonté), par négligence (c.-à-d. sans tenir compte des conséquences de leurs actes par une imprévoyance coupable) ou d'une façon coupable (qui puisse leur être imputée), ni dotées d'une capacité de discernement (c.-à-d. d'une faculté de compréhension subjective, d'une capacité à former une volonté et d'une capacité à agir selon leur volonté).

[46] Cf. 15.3446 Ip. groupe libéral-radical.

Au vu de l'état actuel de la technologie, le Conseil fédéral juge suffisante la réglementation existante. Jusqu'à présent, il n'est pas apparu que son application aux robots crée des lacunes en matière de responsabilité[47]. Cela vaut également pour le droit pénal. En effet, les infractions commises par l'intermédiaire de robots peuvent être poursuivies comme n'importe quel acte commis par un individu au moyen d'un objet. Il n'existe ainsi, en l'état, aucune lacune nécessitant l'intervention du législateur.

Cependant, cette analyse n'exclut pas que la question d'une réglementation spécifique se pose tôt ou tard. Cette situation n'est pas nouvelle : le progrès technique a toujours créé de nouvelles sources de risques, comme les véhicules à moteur, l'énergie nucléaire ou les organismes génétiquement modifiés. Dans tous ces cas, le législateur a réagi avec l'instauration d'une responsabilité à raison du risque. En vertu de cette dernière, un dommage causé par la nouvelle technologie est imputé à une personne en particulier, qui doit ensuite répondre du dommage même sans faute de sa part. Ceux qui profitent de la nouvelle technologie doivent, eux aussi, en supporter les risques[48].

4.3 Explicabilité et transparence

Les décisions fondées sur des systèmes d'IA n'étant souvent pas explicables, des dispositions doivent être prises en vue de garantir la transparence de ces décisions en vertu du respect de l'État de droit.

Une forme d'explicabilité est prévue dans le projet de révision de la loi sur la protection des données : le responsable de traitement (qui peut être une personne privée ou un organe fédéral) doit informer la personne concernée de toute décision qui est prise exclusivement sur la base d'un traitement de données personnelles automatisées, y compris le profilage, et qui a des effets juridiques sur la personne concernée ou qui l'affecte de manière significative (art. 19 du projet[49]). Si la personne concernée le demande, le responsable du traitement privé doit lui donner la possibilité de faire valoir son point de vue. La personne concernée peut exiger que la décision soit revue par une personne physique. En outre, les responsables de traitement devront effectuer une analyse d'impact lorsque le traitement de données personnelles envisagé est susceptible d'entraîner un risque élevé pour la personnalité ou les droits fondamentaux de la personne concernée (art. 20). Des obligations spécifiques s'appliquent aux décisions individuelles automatisées des autorités fédérales.

Lorsque la personne concernée exerce son droit d'accès, le responsable de traitement doit dans tous les cas l'informer de l'existence d'une décision individuelle automatisée ainsi que de la logique sur laquelle se base la décision (art. 23, al. 2, let. f). Le règlement général sur la protection des données fixe des dispositions similaires (cf. notamment art. 15, ch. 1, let. h, RGPD).

Les art. 19 et 23 du projet de révision LPD ne s'appliquent pas lorsqu'il y a une intervention humaine dans la prise de décision et que l'intelligence artificielle constitue simplement une aide à la décision. Par intervention humaine, on entend un examen par une personne physique qui procède à sa propre évaluation de la situation et peut s'écarter du résultat livré par la machine. Un exemple de décision automatisée pourrait être le fait de prononcer une amende pour excès de vitesse exclusivement sur la base de photographies de la plaque minéralogique et de la personne au volant, couplées automatiquement aux données du registre des véhicules et à un outil de reconnaissance faciale.

Des exigences spéciales s'appliquent également à l'explicabilité des décisions individuelles non automatisées prises par les autorités avec l'aide de l'IA et affectant le statut juridique d'une personne. Il découle du droit constitutionnel d'être entendu l'obligation pour les autorités de motiver leurs

[47] Cf. p .ex. 18.3445 Ip. Marchand-Balet concernant les véhicules autonomes et rapport du Conseil fédéral en réponse au postulat Leutenegger Oberholzer 14.4169 « Automobilité » concernant les faits ayant un lien avec l'étranger.

[48] Cf. 17.3040 Po. Reynard.

[49] FF **2017** 7217 ss. ; cf. également 17.059 Loi sur la protection des données. Révision totale et modification d'autres lois fédérales, https://www.parlament.ch/fr/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20170059.

décisions. Si une autorité s'appuie sur l'IA pour prendre une décision, le système doit impérativement fournir des explications concernant les informations et critères dont il a tenu compte, les hypothèses formulées et les motifs déterminants pour le résultat.

Le recours à l'intelligence artificielle par un organe étatique peut nécessiter une base légale qualifiée si des données personnelles sont traitées et que les droits de la personne concernée peuvent s'en trouver gravement atteints. C'est pourquoi le projet de révision LPD prévoit l'exigence d'une base légale formelle lors du traitement de données sensibles ou de profilage ou lorsque la finalité ou le mode de traitement de données personnelles est susceptible de porter gravement atteinte aux droits fondamentaux de la personne concernée (art. 30, al. 2).

Il ressort de ce qui précède que plus l'incidence sur les droits fondamentaux et les droits de l'homme de la personne concernée est importante, plus les exigences relatives à l'explicabilité sont élevées. Par exemple, la recommandation d'un morceau de musique générée par IA n'est en principe pas problématique, tandis que les décisions des systèmes d'IA portant par exemple sur le risque de récidive d'une personne présumée coupable ou condamnée peuvent fortement affecter ses droits fondamentaux.

Les cas où les entreprises utilisent l'IA afin d'interagir avec leurs clients, au moyen de *chatbots*, par exemple, constituent également un défi. Les chatbots peuvent servir de nombreuses manières, notamment à répondre aux questions des clients, à les aider et à les conseiller. Comme il est possible de parler avec un chatbot comme avec un être humain, il peut être difficile à un client de s'apercevoir qu'il parle avec une machine. Selon la recommandation de l'OCDE du 22 mai 2019 sur l'intelligence artificielle, l'utilisation responsable de l'intelligence artificielle suppose d'informer les personnes de l'existence de ces interactions avec des systèmes d'IA. En Suisse, la loi fédérale contre la concurrence déloyale (LCD) pourrait s'appliquer dans les cas où les consommateurs ne sont pas informés au préalable et de façon systématique de l'interaction avec des systèmes d'IA (cf. chapitre 6.15).

4.4 Biais et discrimination

Les erreurs systématiques dans les données ou les algorithmes des systèmes d'IA peuvent entraîner la discrimination systématique de certaines catégories de personnes par les applications. Notamment, si les données comportent des biais discriminatoires, il y a de fortes chances que ces biais soient reproduits par le système algorithmique. Certes, ce n'est pas une caractéristique spécifique à l'IA, puisque toutes les données ou décisions reposant sur des algorithmes peuvent, dans la mesure où ils se fondent sur des décisions humaines antérieures, reproduire et multiplier les mêmes préjugés susceptibles de fausser la prise de décision humaine. Avec le phénomène de la boîte noire et la possibilité d'automatiser la procédure de décision, le problème peut toutefois être considérablement accentué.

En outre, les possibilités nouvelles permettant d'analyser d'importantes quantités de données peuvent entraîner l'apparition de problèmes inédits avec l'utilisation de l'IA. Ainsi peut-il être non souhaitable ou interdit par la loi de prendre en compte certaines informations dans les décisions. Dans certains cas, un système IA peut néanmoins identifier ces informations dans des données en apparence indépendantes ou neutres (données indirectes). Dans le cadre d'une procédure de recrutement, par exemple, la question au sujet de la grossesse d'une candidate pourrait ne pas être posée, mais l'information pourrait être obtenue à partir d'autres données[50].

Les prédictions algorithmiques peuvent même amplifier la discrimination : par exemple, si la police a arrêté davantage de personnes issues de la migration par le passé et qu'elle a concentré son attention sur certains quartiers où résident davantage de personnes issues de la migration, il y a de

[50] L'étude du Conseil de l'Europe « Discrimination, artificial intelligence and algorithmic decision-making » cite un exemple où l'existence d'une grossesse a pu être détectée à partir des achats effectués par la personne concernée.

fortes chances que ce dernier reproduise ces biais. Il en résultera davantage d'arrestations que dans d'autres quartiers ou catégories de population qui font l'objet d'une attention moindre, ce qui renforcera encore l'idée reçue selon laquelle ces catégories de personnes et ces quartiers posent davantage de risques en matière de criminalité (« feedback loop »).

Il n'est pas non plus exclu que le recours à l'intelligence artificielle crée de nouvelles catégories de discriminations. Si par exemple un système de prédiction se base sur le fait que 80% des résidents d'un quartier déterminé paient leurs factures avec retard et qu'une compagnie se base sur ces résultats pour refuser tout crédit aux habitants de ce quartier, elle désavantagera également les 20% de résidents qui paient leurs factures à temps.

L'étude du Conseil de l'Europe « Algorithmes et droits humains »[51] recommande, pour savoir si un algorithme favorise ou au contraire permet de prévenir un traitement discriminatoire, de se référer à la distinction faite sur le plan juridique entre discrimination directe et discrimination indirecte. Cette distinction est connue en droit suisse, en particulier dans la loi sur l'égalité[52]. On parle de discrimination directe lorsqu'une personne fonde sa décision sur l'appartenance d'une personne à un groupe déterminé qui est ou fut l'objet de discriminations par le passé. On parle de discrimination indirecte lorsque la décision se fonde sur des facteurs d'apparence neutre mais qui ont pour effet de désavantager certains groupes de population qui font ou firent l'objet par le passé de discriminations (par exemple le fait de travailler à temps partiel pour désavantager les femmes).

Le droit suisse permet d'appréhender ces risques dans une certaine mesure, y compris dans les rapports entre particuliers. Les lois existantes s'appliquent aussi dans le cas de discriminations résultant de systèmes d'IA. La loi sur l'égalité, qui interdit les discriminations fondées sur le sexe à l'embauche et dans les rapports de travail de droit privé et de droit public, s'appliquerait par exemple à un processus de sélection fondé sur l'intelligence artificielle.

Les mesures prévues par la révision LPD (voir ch. 4.3 ci-dessus) peuvent avoir un effet préventif et correctif : par exemple, la personne concernée a la possibilité de faire valoir son point de vue en cas de décision individuelle automatisée et elle peut exiger que la décision soit revue par une personne physique.

Une attention particulière devrait être portée sur les risques de discriminations lors du recours à l'intelligence artificielle dans le secteur public, que ce soit pour élaborer des normes de droitou pour appliquer le droit. Cela peut passer par une analyse de risque lors de la conception des systèmes et par un monitorage et des évaluations régulières.

Il convient de garder à l'esprit que le risque de discrimination n'est pas le même si l'on recourt à l'intelligence artificielle dans un jeu d'échec ou pour sélectionner des candidats à l'embauche. Enfin, il ne faut pas oublier que, si elle est bien faite, l'intelligence artificielle peut aussi identifier et éviter les biais humains, et donc contribuer à réduire les décisions discriminatoires.

4.5 Accès aux données et protection des données

L'accès aux données et leur disponibilité sont essentiels, en particulier pour la recherche scientifique, mais aussi pour les applications pratiques de l'IA dans l'économie et la société. À cet effet, la Confédération dispose d'une politique des données. Celle-ci vise en premier lieu à favoriser l'accès aux données, notamment à des données librement accessibles (Open Data) en tant que matière première d'une économie et d'une société numériques, ainsi qu'à instaurer des bases légales et des conditions-cadre cohérentes et adaptées à l'époque permettant à la Suisse de se positionner comme un pôle attractif en matière de création de valeur par les données. Mais la politique des données

[51] *Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires*, étude du Conseil de l'Europe DGI (2017)12, p. 29 ss.

[52] RS **151.1**

définit également le cadre juridique à l'intérieur duquel les données peuvent être recueillies, liées et analysées de manière licite.

Dans le même temps, la Confédération doit garantir la sécurité et la protection des données. Cela touche, d'une part, à la protection des données personnelles dans le cadre de la politique de protection des données et, d'autre part, à la protection de la propriété intellectuelle, qui peut être concernée lors du traitement ou de l'utilisation de données, par exemple lorsque des textes ou des images protégés par le droit d'auteur doivent être mis à disposition pour des applications IA.

Si les atteintes à la protection des données ne sont pas le seul fait des IA, celles-ci peuvent toutefois constituer un facteur aggravant. De nos jours, avec la multiplication des empreintes numériques, les algorithmes permettent d'établir des profils détaillés de personnes sans avoir à interroger ces dernières. La possibilité de combiner des données personnelles issues de différentes sources, en particulier, confère aux applications d'IA un potentiel inédit.

En **droit de la propriété intellectuelle**, il s'agit d'une part de droits de propriété adéquats pour les systèmes d'IA exercés à travailler partiellement avec des données soumises à des dispositions légales. D'autre part, les systèmes d'IA peuvent aussi « créer » du neuf, ce qui pose la question de savoir si les œuvres littéraires ou artistiques et les inventions nées de l'utilisation d'IA sont protégées par le droit d'auteur ou le droit des brevets et, le cas échéant, à qui reviennent ces droits[53].

En **droit des brevets**, l'opinion généralement admise en Suisse est que seules les personnes physiques peuvent inventer quelque chose au sens juridique – voire, selon les interprétations, des personnes morales. Les systèmes d'IA sont exclus de ce champ, du moins pour l'instant, en raison du vide juridique qui les entoure.

Il convient par ailleurs de clarifier comment le **droit d'auteur** doit s'appliquer si de nombreuses formes d'IA utilisent d'énormes quantités de données dans leur processus d'apprentissage et que ces données sont en partie protégées par le droit d'auteur. Les données doivent être en général reproduites de nombreuses fois par les IA, ce qui constitue en principe une violation du droit d'auteur. Cela pourrait présenter un obstacle de taille si l'on veut développer des IA.

La réglementation dans ces domaines doit par conséquent tenir compte des conflits d'objectifs entre ces différents droits. Les défis généraux relevant de ce domaine thématique sont abordés au chapitre 0 du présent rapport. Par ailleurs, d'autres questions spécifiques peuvent se poser concernant l'accès aux données et leur sécurité. Elles sont traitées dans les sections portant sur les domaines thématiques correspondants.

5 Intelligence artificielle – recherche, développement et application en Suisse

Cette section propose une brève analyse de la position de la Suisse en matière d'intelligence artificielle (IA). Après une présentation des principaux acteurs du domaine (paysage suisse de la recherche en IA), trois thèmes sont analysés au moyen d'indicateurs quantitatifs: comment se positionne la Suisse en matière de recherche, de développement et d'application de l'intelligence artificielle[54] ?

[53] TA-SWISS (éd.). *Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz*, manuscrit non publié.

[54] Il convient de relever que la mesure du développement des technologies liées à l'IA est délicate : les frontières entre l'IA et les autres technologies sont floues et en constante évolution.

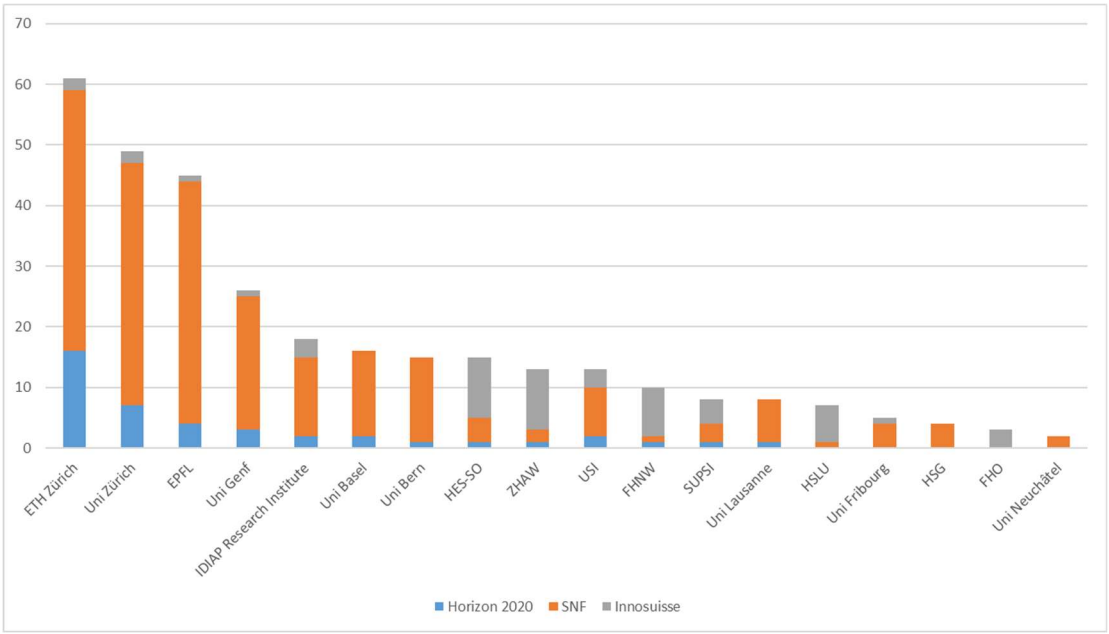
5.1 Paysage suisse de la recherche en IA : vue d'ensemble des acteurs

La Suisse est devenue un pôle important dans le domaine de l'IA. Varié et dynamique, le paysage de la recherche comprend des centres de recherche renommés et établis de longue date, parmi lesquels l'Istituto Dalle Molle di Studi sull'Intelligenza Artificiale IDSIA (intégré à USI et à la SUPSI dans la figure 5) et l'IDIAP Research Institute (anciennement Institut d'intelligence artificielle perceptive), ainsi que les centres du Domaine des EPF. En outre, le Domaine des EPF développe actuellement son grand axe stratégique Sciences des données, tandis que l'EPFL et l'ETH Zurich ont créé conjointement le Swiss Data Science Center (SDSC). Les hautes écoles spécialisées et les autres universités sont, elles aussi, actives dans le domaine de l'IA. Le Datalab, un regroupement virtuel de plusieurs départements de la ZHAW, travaille par exemple en étroite collaboration avec l'industrie. Pour les entreprises opérant en Suisse dans le secteur de l'IA, le niveau élevé de la recherche joue un rôle crucial. Des initiatives privées telles que le Swiss Group of Artificial Intelligence and Cognitive Science (SGAICO) complètent les initiatives du paysage des hautes écoles en rassemblant chercheurs et utilisateurs et promeuvent la transmission des connaissances, l'instauration d'un climat de confiance et l'interdisciplinarité.

Même la promotion de la recherche s'est emparée de l'IA. À travers le FNS, la Confédération investit par exemple dans le PNR 75 « Big Data », le PNR 77 « Transformation numérique » et le NCCR « Robotique ». Via Innosuisse, elle soutient le réseau thématique national Swiss Alliance for Data-Intensive Services (Data + Service), qui accompagne les entreprises dans le développement de nouveaux produits et services, encourageant ainsi le transfert de connaissances, tout comme dans le programme d'impulsion « technologies de fabrication », où l'IA est également présente.

Les bases de données du FNS, d'Innosuisse et de l'UE (Horizon 2020) permettent de dresser une vue d'ensemble sommaire des instituts de recherche suisses actifs dans le domaine. Comme le montre la Figure 5, le nombre de **projets** encouragés par le FNS, par Innosuisse et dans le cadre du programme Horizon 2020 indique que la recherche suisse en IA est très diversifiée et répartie dans l'ensemble des régions du pays. Outre les deux EPF, les universités et les hautes écoles spécialisées sont également actives dans la recherche sur l'IA. Ces dernières s'engagent principalement dans le domaine du TST, contribuant de ce fait largement à la diffusion des connaissances dans les milieux économiques.

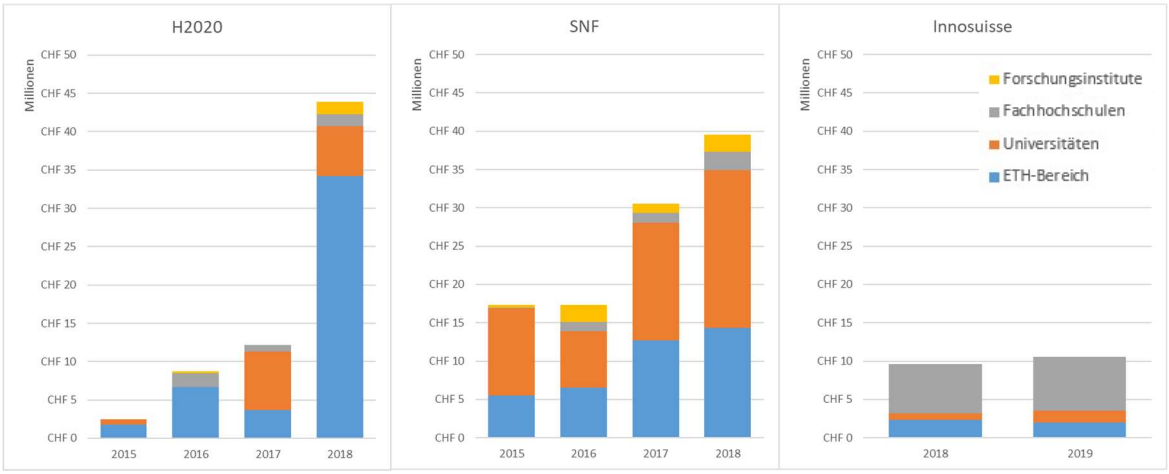
Figure 5 : Promotion de projets IA (Horizon 2020, FNS et Innosuisse) : nombre de projets 2015-2018



Source : Analyse SEFRI à partir de bases de données du FNS (P3), d'Innosuisse (2018 – 2e trimestre 2019) et de l'UE ; les recherches menées dans les bases de données se sont appuyées sur les mêmes mots-clés relatifs à l'IA.

L'examen des **volumes d'encouragement** révèle une nette augmentation au cours des dernières années (Figure 6), notamment pour ce qui concerne la recherche soutenue par l'UE. Une tendance claire émerge de la répartition des subventions. Alors que le Domaine des EPF domine les projets européens, l'encouragement par le FNS concerne majoritairement les universités. Dans le domaine de l'IA, Innosuisse collabore principalement avec les hautes écoles spécialisées, avec néanmoins des volumes d'encouragement globalement beaucoup plus faibles.

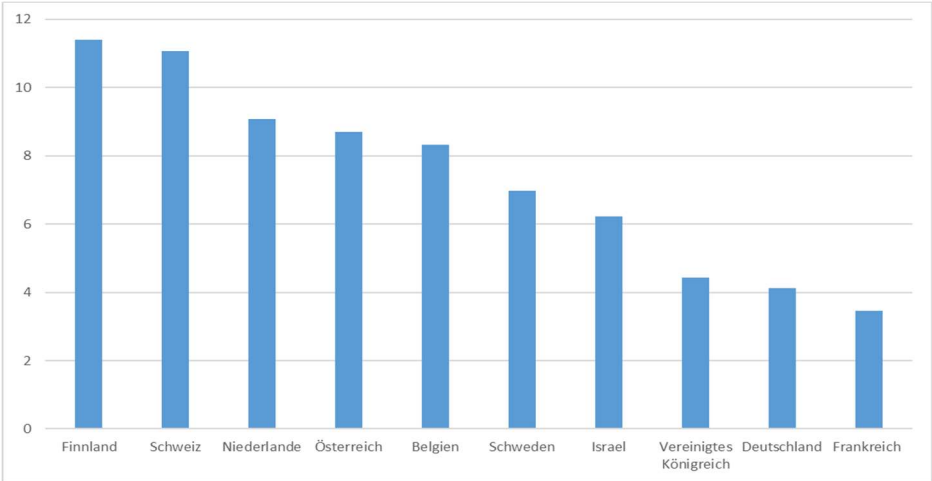
Figure 6 : Évolution des volumes d'encouragement de la recherche sur l'IA



Source : Analyse du SEFRI à partir de bases de données du FNS (P3), d'Innosuisse (2018 – 2e trimestre 2019) et de l'UE.

La comparaison internationale dans le cadre du programme Horizon 2020 montre que la Suisse se positionne bien si l'on considère le nombre de projets IA par million d'habitants (figure 7). Cela peut suggérer que le paysage suisse de la recherche en IA est compétitif au plan international.

Figure 7 : Nombre de projets IA par million d'habitants, Horizon 2020, pays sélectionnés, 2015-2018



Source : Analyse du SEFRI à partir de bases de données de l'UE.

5.2 Performance de la R-D en Suisse

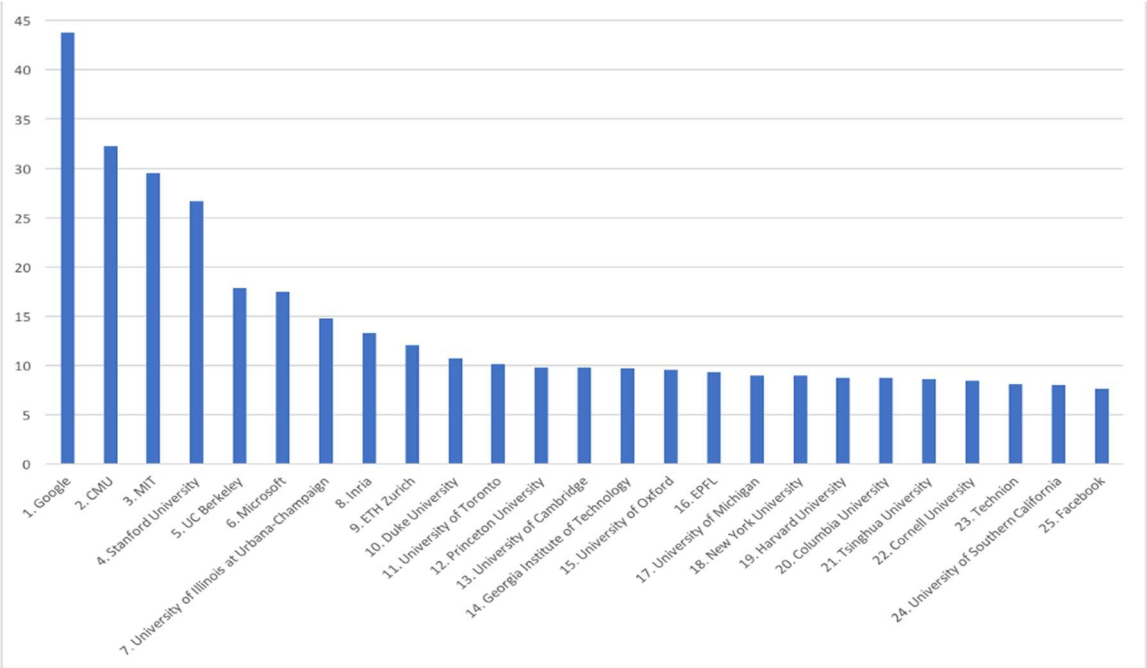
5.2.1 Recherche en IA en Suisse

Les écoles polytechniques fédérales se classent aux huitième et neuvième places du classement international QS World University Ranking 2019 dans la catégorie des écoles informatiques (sans

illustration). Outre les coopérations de recherche avec les universités et les entreprises suisses, ce sont les contacts internationaux qui contribuent à cet excellent résultat. Par exemple, l'Institute for Machine Learning de l'ETH Zurich collabore avec des universités américaines prestigieuses, l'EPFL et le « Max-Planck-Institut für intelligente Systeme ». Non seulement les groupes technologiques comme Google, Facebook, IBM ou Microsoft coopèrent étroitement avec la recherche suisse, mais ils embauchent aussi massivement les talents issus des hautes écoles. Outre l'ETHZ et l'EPFL, d'autres universités suisses sont très performantes dans les sciences informatiques, ce qui explique que les entreprises recrutent des collaborateurs de haut niveau et engagent des coopérations de recherche avec des laboratoires universitaires.

Dans le domaine de l'IA, la qualité des établissements peut être mesurée au nombre de publications à la « Conference on Neural Information Processing Systems » (NeurIPS), la conférence la plus renommée en la matière. Là encore, cet indicateur montre que les deux EPF comptent parmi les hautes écoles les plus réputées du monde (Figure 8).

Figure 8 : Nombre de publications à la « Conference on Neural Information Processing Systems » 2017 par organisation



Source : Conference on Neural Information Processing Systems (NeurIPS), voir également https://medium.com/@chuvpilo/whos-ahead-in-ai-research-insights-from-nips-most-prestigious-ai-conference-df2c361236f6

Pour les chercheurs, la publication d'articles dans des journaux scientifiques est le principal moyen de diffusion des connaissances. L'analyse statistique des publications scientifiques fournit donc un bon aperçu des activités de recherche d'un pays ou d'une région du monde.

Si l'on examine le volume des publications scientifiques suisses dans le domaine de l'IA, on constate que la recherche est dominée par les grands pays en termes quantitatifs mais que **la recherche suisse se distingue en termes qualitatifs.** Selon le Times Higher Education, la Chine a produit plus de 41 000 publications dans le domaine de l'IA sur la période 2011-2015, soit plus du double des Etats-Unis (25 500). Le Japon (11 700) prend la troisième place, le Royaume-Uni (10 100) la quatrième et l'Allemagne (8 000) la cinquième. La Suisse enregistre 1 700 publications dans le domaine de l'IA sur la période 2011-2015. Proportionnellement au nombre d'habitants, il s'agit d'un chiffre élevé mais inférieur à celui de Singapour, Hong Kong ou les Pays-Bas (sans illustration). Toutefois, si la Chine obtient un score élevé selon le critère du volume, elle ne se classe que 34e sur le plan **de l'impact des citations** pondéré par discipline. Selon cet indicateur, qui mesure plutôt la qualité et la réception des publications au sein de la communauté scientifique, la Suisse occupe la

première place mondiale avec un facteur d'impact des citations de 2,71, suivie de Singapour (2,24) et de Hong Kong (2,00).

Tableau 3 : Volume et impact des publications dans le domaine de l'IA par pays (2011-2015)

Classement	Pays	Impact des citations pondéré par discipline (1 correspond à la moyenne)	Nombre de publications
1	Suisse	2,71	1685
2	Singapour	2,24	2432
3	Hong Kong	2,00	2205
4	États-Unis	1,79	25 471
5	Italie	1,74	6221
6	Pays-Bas	1,71	2458
7	Australie	1,69	5227
8	Allemagne	1,66	7957
9	Belgique	1,64	1537
10	Royaume-Uni	1,63	10 120

Remarques : le tableau présente uniquement les dix pays ayant le plus grand impact.
Source : Rapport sur l'IA d'Elsevier.

La mise au point de techniques d'apprentissage automatique (machine learning) a joué un rôle déterminant dans la recherche dans le domaine de l'IA. On observe depuis quelques années un accroissement notable des **publications scientifiques liées à l'apprentissage automatique**. Les Etats-Unis sont en tête des efforts de recherche dans ce domaine, avec un nombre de citations de publications 3,3 fois plus élevé que celui de la Chine, deuxième du classement, et 5,6 fois plus élevé que celui du Royaume-Uni, troisième du classement. Dans le domaine de l'apprentissage automatique, la Suisse occupe le 14e rang de ce classement, se situant ainsi au même niveau que des pays comme Singapour ou les Pays-Bas. Ses bons résultats dans ces classements internationaux confirment l'excellente qualité de la recherche suisse en matière d'IA.

5.2.2 Développement dans le domaine de l'IA en Suisse (analyse des brevets)

Les demandes de brevets reflètent l'exploitation technologique et commerciale des connaissances issues de la recherche. En ce sens, elles peuvent être considérées comme un indicateur des activités de développement.

Selon l'OCDE, à eux trois le Japon, la Corée du Sud et les Etats-Unis produisent plus de 60% des brevets déposés dans le domaine de l'IA. Avec 10,4%, la Chine occupe le quatrième rang. Tous les autres pays sont en-dessous de la barre des 5%. La Suisse se classe **en 16e position avec 0,4%** du total mondial des brevets déposés dans le domaine de l'IA.

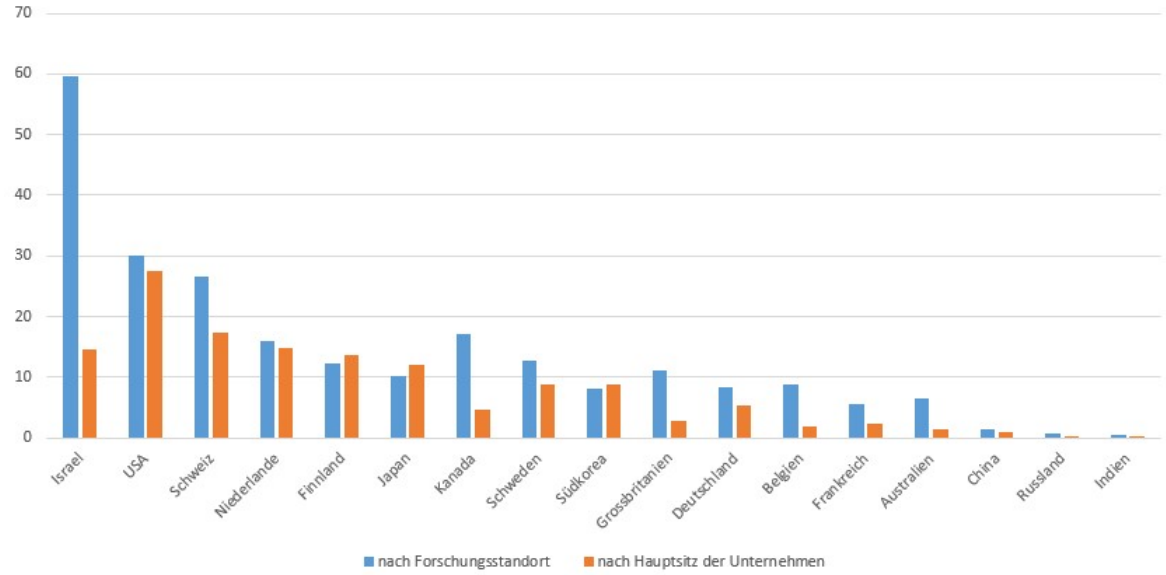
Si l'on se concentre sur les 2000 entreprises les plus actives en R-D (qui, regroupées, détiennent 75% des inventions protégées dans les cinq grandes institutions de la propriété intellectuelle), on constate que celles qui sont installées au Japon, en Corée, au Taipei chinois et en Chine possèdent environ 70% de l'ensemble des brevets déposés dans le domaine de l'IA. La part des entreprises basées aux Etats-Unis s'élève quant à elle à 18% (3e rang). La Suisse occupe **la 13e position**. La légère amélioration du classement de la Suisse par rapport à l'indicateur précédent pourrait être due au fait qu'elle abrite le siège de nombreuses grandes entreprises ; or, c'est souvent depuis leur siège que les multinationales déposent leurs demandes de brevets[55].

[55] OCDE (2017) : Science, technologie et industrie : Tableau de bord 2017 – La transformation numérique, disponible à l'adresse https://www.oecd.org/fr/sti/science-technologie-et-industrie-tableau-de-bord-de-l-ocde-20747217.htm

Si l'on base la comparaison non pas sur les chiffres absolus, mais sur le **nombre de brevets par million d'habitants**, le classement de la Suisse au niveau mondial est bien meilleur. De même, l'analyse de la qualité des brevets modifie aussi la position de la Suisse : si la Chine, par exemple, dépose un très grand nombre de brevets, leur degré d'innovation technique et leur applicabilité économique sont en revanche nettement plus faibles que ceux publiés dans beaucoup d'économies avancées, y compris la Suisse.

Le graphique ci-dessous présente le nombre de brevets de classe mondiale[56] déposés dans le domaine de l'IA par rapport à la population selon le site de recherche ou le siège des entreprises. La Suisse occupe la **3e place** derrière Israël et les États-Unis.

Figure 9 : Nombre de brevets de classe mondiale dans le domaine de l'IA par million d'habitants en 2018



Source : SEFRI d'après EconSight : « Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung », 2019.

Au vu du niveau de la recherche fondamentale et appliquée sur l'IA, la Suisse dispose donc globalement d'une **base solide pour réaliser le transfert de savoir et de technologie (TST)**.

5.2.3 Application de l'IA en Suisse

Les start-up fournissent des éléments permettant d'éclairer le dynamisme des applications d'IA innovantes. Il apparaît par exemple que les États-Unis créent un très grand nombre de start-up spécialisées dans l'IA en comparaison internationale. Les États-Unis réussissent remarquablement à transférer les résultats issus de la recherche fondamentale vers l'économie par le biais des créations d'entreprises. Selon une étude de Roland Berger, près de 40 % des start-up spécialisées dans l'IA sont situées aux États-Unis[57].

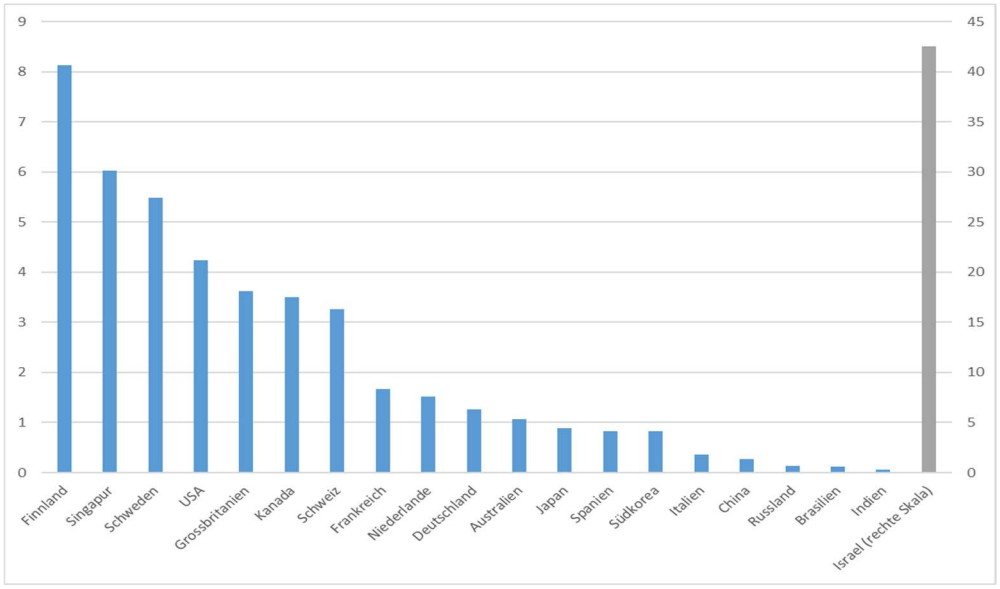
L'Europe se classe derrière les États-Unis, tandis que la Suisse se situe dans la moyenne des pays considérés. Si l'on tient compte de la taille de la population, la Suisse se place dans le peloton de tête en nombre de start-up (figure 10). Mais comme l'étude de Roland Berger tire ses analyses de plusieurs sources nationales, les chiffres ne sont comparables que dans une certaine mesure (il se

[56] Les brevets de classe mondiale sont les 10 % des brevets qui, pour chaque technologie, sont les mieux classés au niveau mondial, en fonction de la couverture du marché et des citations dans d'autres brevets.

[57] Cf. Roland Berger. *Artificial Intelligence, A strategy for European startups*, 2018. https://www.rolandberger.com/publications/publication_pdf/roland_berger_ai_strategy_for_european_startups.pdf

peut par exemple que le nombre élevé concernant Israël soit surévalué en comparaison internationale).

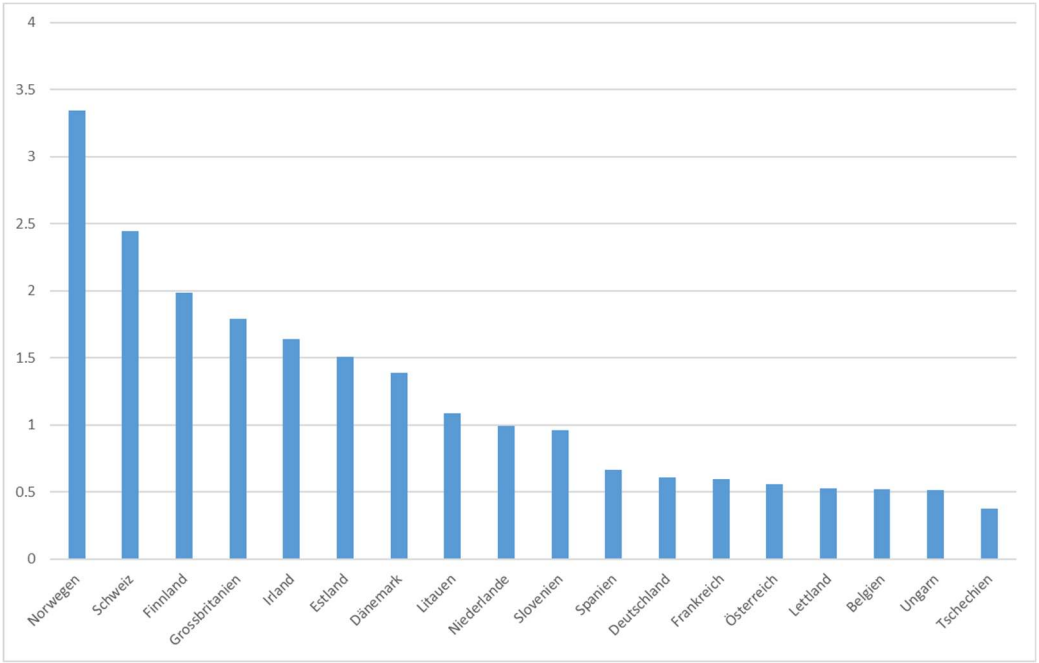
Figure 10 : Nombre de start-up spécialisées dans l'IA dans le monde par million d'habitants en 2018



Source : calculs du SEFRI d'après Roland Berger : « Artificial Intelligence: A strategy for European startups », 2018.

D'après une autre étude réalisée par ASGARD en 2017 et qui porte uniquement sur les pays européens, la Suisse se situe parmi les pays les plus dynamiques en termes de création de start-up actives dans le domaine de l'IA (figure 11) et peut prétendre à un rôle pionnier en comparaison avec les autres pays d'Europe – et ce, aussi bien en valeur absolue qu'au prorata du nombre d'habitants.

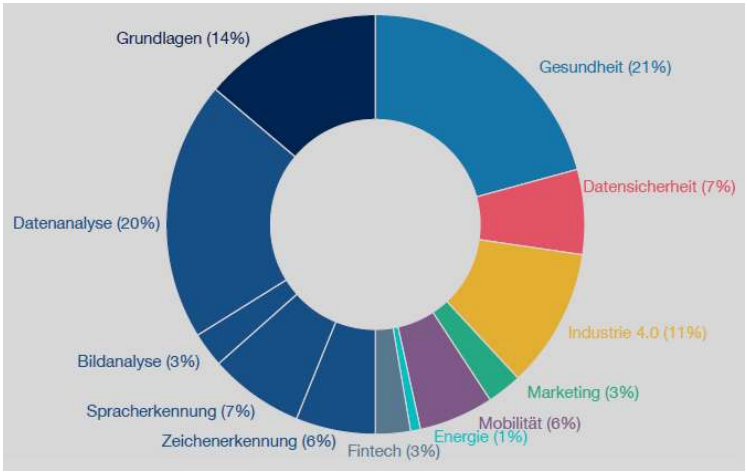
Figure 11 : Nombre de start-up spécialisées dans l'IA en Europe par million d'habitants en 2018



Source : calculs du SEFRI d'après ASGARD. « The European Artificial Intelligence Landscape », 2017.

L'IA trouve des applications dans presque tous les domaines et les branches. On peut notamment citer les transports, la santé, la sécurité (numérique et physique), la finance, le marketing ou encore l'éducation. Toutefois, si l'on prend en considération les axes prioritaires de développement, l'étude d'EconSight montre que la Suisse se caractérise par une prépondérance des applications de l'IA dans[58]. Le dépôt de brevets dans ce domaine est d'ailleurs en forte hausse depuis une dizaine d'années. L'analyse de données et l'industrie 4.0 constituent deux **autres domaines de poids dans le portefeuille de brevets suisses**. Si l'on examine les brevets sous l'angle de la structure économique, il existe donc quelques branches importantes très demandeuses d'IA (industrie pharmaceutique, technologies médicales et autres industries) et d'autres (comme la finance) dans lesquelles les brevets d'IA sont en moyenne sous-représentés.

Figure 12 : Structure de l'IA en Suisse, sous-technologies et domaines d'application, 2018



Source : EconSight. « Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung », 2019.

5.3 Défis dans le domaine de la recherche et de l'innovation

Même si, comme montré dans les paragraphes précédents, la Suisse est dans une situation favorable pour ce qui concerne la qualité de la recherche et le dynamisme de l'innovation dans le domaine de l'IA, les défis restent importants. Le changement structurel que nous vivons actuellement est inédit de par son **envergure** et la **vitesse à laquelle les technologies numériques pénètrent** les différents secteurs. Ce faisant, les technologies numériques transforment aussi de plus en plus de technologies clés dans d'autres secteurs – y compris des domaines dans lesquels la Suisse est à la pointe. La recherche et la formation jouent un rôle crucial dans la maîtrise de ces défis. De plus, le transfert de savoir et de technologie (TST) est décisif pour permettre une mise en œuvre rapide des résultats de la recherche fondamentale et ainsi garantir la compétitivité dans toutes les branches.

Dans le contexte de la numérisation, les technologies de l'IA tiennent une **place particulière**. Elles permettent l'automatisation de tâches qui nécessitaient auparavant l'aptitude à la perception, à la déduction et à l'interaction propre à l'homme. Le développement et la propagation de l'IA ne sont pas isolés des autres développements technologiques, mais font partie intégrante du **processus général de numérisation**. Outre les progrès dans le domaine des logiciels ou de l'IA, les principaux moteurs de la numérisation sont les avancées en matière de robotique, de technologie des capteurs et de fabrication additive. La transformation numérique est également portée par les progrès dans les technologies des processus et de la mémoire ou encore le réseautage accru de l'information. Ces moteurs interagissent étroitement et ne peuvent pas être dissociés. C'est pourquoi la Confédération

[58] Cf. EconSight. *Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung*, 2019. https://www.econsight.ch/artificial-intelligence/

adopte une **vision d'ensemble** de la numérisation en matière de politique de la recherche et de l'innovation.

Dans cette perspective large, des **faiblesses ponctuelles** existent dans plusieurs domaines de la numérisation, y compris la recherche[59]. En comparaison internationale, compte tenu notamment des capacités de recherche nécessaires pour couvrir au plus haut niveau l'ensemble de la numérisation, la Suisse est à la traîne au sein des champs de recherche dédiés aux TIC dans des domaines de recherche importants. Ce constat vaut également pour l'acquisition, le traitement, le stockage, la gestion et la diffusion d'informations (éléments clés du *Big Data*) et pour l'utilisation des technologies numériques dans la communication entre systèmes et appareils (aspects centraux de « l'Internet des objets » et de « l'industrie 4.0 »). Dans ces domaines, l'augmentation de la prestation de recherche en Suisse comparée à celle des pays les plus performants est nettement inférieure à la moyenne, de sorte que la Suisse a perdu beaucoup de terrain.

Dans ce contexte, le DEFR a étudié en 2017 les défis de la numérisation pour la formation et la recherche en Suisse et a lancé le plan d'action « Numérisation pour le domaine FRI durant les années 2019 et 2020 ». Le plan d'action vise à renforcer les compétences dans les domaines de la formation et de la recherche. La plupart des mesures définies pour les différents champs d'action ont déjà été mises en œuvre ou sont en cours de mise en œuvre.

Eu égard au rythme fulgurant des développements technologiques, le plan d'action prévoyait que plusieurs activités soient lancées rapidement dans des domaines spécifiques et menées en étroite collaboration avec les acteurs concernés. Toutefois, le système suisse, qui accorde une grande importance à l'autonomie des acteurs, ayant parfaitement fait ses preuves, le financement prévu dans le cadre du plan d'action a pris la forme d'un financement incitatif limité dans le temps. Les mesures ont donc été consolidées au cours de la période FRI 2021-2024 et doivent être mises en œuvre par les acteurs de manière autonome.

De nombreux pays industrialisés, dont le Canada et la Finlande, ont qualifié l'IA de technologie d'avenir de premier plan. Cependant, les **stratégies nationales de promotion de l'IA sont très variables[60]**. Il s'avère que chaque pays définit ses propres axes prioritaires dans le but de promouvoir l'IA dans le secteur public et privé. Par exemple, la Chine entend devenir le numéro un de l'IA d'ici 2030. En débloquant des ressources financières considérables en faveur de la recherche et du développement, elle souhaite réduire son écart avec les États-Unis, qui sont actuellement leaders du marché de l'IA. À l'inverse, la Finlande met l'accent sur l'application de l'IA et la coopération entre le secteur économique, le monde scientifique et l'administration.

La Suisse dispose d'une politique de l'innovation qu'elle ne met pas en œuvre dans le cadre d'une « stratégie d'innovation » unique et globale, mais de façon décentralisée, au sein de plusieurs domaines politiques à la fois autonomes et coordonnés par thème. Ce mode d'organisation octroie une grande marge de manœuvre aux acteurs et leur permet de réagir de façon diversifiée et adaptée aux défis et aux opportunités qui se présentent, comme le montre le thème de la numérisation[61]. Cette approche s'applique également à l'IA en tant que sous-domaine de la numérisation.

Bien que la Suisse dispose d'excellents établissements de recherche, engage des moyens substantiels dans le domaine de la numérisation et propose des instruments efficaces de promotion de la R-D dans les technologies numériques, le développement des compétences dans le domaine FRI doit être renforcé compte tenu de la forte dynamique technologique et de la concurrence croissante. Pour ce faire, les activités liées à l'IA doivent être intensifiées dans le domaine FRI. Ce sont avant tout les acteurs publics et privés qui doivent s'y atteler de manière autonome, mais cela

[59] Cf. SEFRI. *Défis de la numérisation pour la formation et la recherche en Suisse*, 2017. https://www.sbfi.admin.ch/dam/sbfi/fr/dokumente/webshop/2017/bericht-digitalisierung.pdf.download.pdf/bericht_digitalisierung_f.pdf

[60] OCDE. « Artificial Intelligence in Society », 2019.

[61] Voir aussi : Conseil fédéral. *Vision d'ensemble de la politique d'innovation*, rapport rédigé en réponse au postulat Derder 13.3073 du 13 mars 2013, 2018.

doit aussi passer par le recours aux instruments de promotion existants de la Confédération en faveur du développement des compétences et du TST.

Dans le cadre des structures et des organes existants dédiés à la numérisation, la politique FRI doit donc garantir de manière plus marquée que les acteurs du monde scientifique, de la formation et du TST sont préparés aux défis de la numérisation et relèvent ceux de l’intelligence artificielle dans le cadre de leurs activités respectives en matière de numérisation.

6 Domaines thématiques IA par domaines politiques

Le principal mandat confié au groupe de travail interdépartemental Intelligence artificielle (GTI IA) par le Conseil fédéral était d’assurer l’échange de connaissances et de vues et de coordonner les positions de la Suisse au sein des instances internationales. C’est pourquoi tous les départements étaient représentés dans le groupe de travail mis en place par le DEFR (SEFRI).

Dans le cadre de son mandat, le GTI IA a effectué une analyse des défis liés à l’IA touchant la Confédération. Il a identifié 17 domaines thématiques importants devant être prioritairement examinés selon la Confédération.

Ces domaines thématiques ont été traités sous la responsabilité de l’office compétent. Les défis liés aux applications d’IA étant très variables selon le domaine thématique, les consultations nécessaires ont été plus ou moins importantes. Alors que certains domaines thématiques étaient déjà suffisamment traités par les offices, des groupes de projet représentatifs ont dû être mis en place pour d’autres.

Au total, sept grands groupes de travail interdépartementaux ont été créés, et de nombreux experts et parties prenantes externes du monde scientifique et économique ont été consultés. Conformément au mandat confié par le Conseil fédéral, les réflexions concernant une utilisation transparente et responsable de l’intelligence artificielle ont été intégrées aux travaux.

Les conclusions des travaux des offices compétents ou des groupes de travail sont résumées ci-après. Toutes les sections suivent une structure identique :

- Dans une première section, l’**importance** de l’utilisation de l’IA est expliquée pour chaque domaine thématique.
- Dans une deuxième section, les **défis spécifiques** concernant l’IA sont décrits, en particulier dans le domaine de compétence de la Confédération.
- La présentation des activités existantes expose ensuite les **mesures** que la Confédération ou les acteurs externes concernés ont déjà engagées pour relever ces défis.
- Enfin, le rapport **évalue** si les défis identifiés dans les différents domaines thématiques sont suffisamment traités au moyen des activités existantes (ou de la réglementation existante). Dans la négative, le rapport propose une première analyse des **actions supplémentaires** requises le cas échéant pour y répondre en temps utile.

6.1 Instances internationales et intelligence artificielle[62]

6.1.1 Vue d'ensemble

La majeure partie des développements et défis actuels liés à l'IA revêtent une dimension internationale. En conséquence, l'IA est aujourd'hui un thème central très débattu sur la scène internationale. D'une part, l'intérêt à l'égard de la coopération internationale se fonde sur la nécessité de mettre en commun les ressources destinées à la recherche et au développement et d'assurer l'accès à des sources de données importantes (*Big Data*). D'autre part, le monde numérisé est de plus en plus mobile, et les données, produits et services franchissent les frontières nationales. Aussi les appels à l'instauration de principes éthiques et de normes internationales se multiplient-ils afin de permettre une exploitation optimale du potentiel positif de l'IA, tout en identifiant et en combattant les risques afférents. De plus en plus d'organisations internationales (p. ex. l'ONU, l'UNESCO, l'OCDE, le Conseil de l'Europe et l'UE) et d'instances techniques telles que l'IEEE Standards Association s'intéressent aux questions soulevées par l'utilisation de l'IA. Les discussions portent notamment sur les conditions-cadre et les modèles de gouvernance qu'il faudrait instituer au niveau international pour que la transparence et la compréhension de l'IA soient garanties, les valeurs éthiques fondamentales respectées et les responsabilités clarifiées et afin que le changement structurel soit supportable pour nos sociétés. À cet effet, il convient de s'appuyer sur les règles et normes existantes, par exemple en matière de droits de l'homme, de protection des données et de responsabilité sociale des entreprises.

6.1.2 Défis

Les processus de réglementation multilatéraux traditionnels sont souvent lourds et ont du mal à suivre le rythme des développements que connaît l'IA. En raison des divergences affichées par les pays membres au sujet du rôle des États, de nombreuses institutions de l'ONU subissent de fréquents blocages sur la question de la gouvernance numérique mondiale. De surcroît, un petit nombre de sociétés technologiques internationales possèdent des ressources en données et financières considérables, ce qui les met en position de renforcer leur pouvoir sur le marché également dans le domaine de l'IA. La marge de manœuvre des gouvernements nationaux et des organisations internationales est de plus en plus remise en cause. Pour les petits États comme la Suisse, il est difficile dans ce contexte d'imposer leurs réglementations nationales dans le reste du monde. Le rôle de la coopération internationale devient donc primordial.

Au cours des dernières années, quelques entreprises leaders du secteur de l'IA ont élaboré leurs propres principes en la matière. Dans les débats internationaux, on entend cependant un nombre grandissant de voix affirmer que l'autorégulation de l'industrie ne suffit pas (plus) pour garantir une utilisation transparente, compréhensible et responsable de l'IA. De plus, les processus d'élaboration de normes éthiques des sociétés technologiques transnationales souffrent d'un déficit démocratique. Dans ce contexte, l'intérêt pour les nouvelles approches de réglementation « intelligentes » ne cesse de croître. Les structures de gouvernance dynamiques, flexibles, interdisciplinaires et décentralisées, de même que la création de normes internationales jouissant d'une légitimité démocratique, sont de plus en plus plébiscitées. Du point de vue de la Suisse, il est notamment primordial d'impliquer toutes les parties prenantes concernées du monde entier – qui, outre les États, comprennent aussi le secteur privé, la société civile et les experts techniques – dans les processus de décision politiques et de les responsabiliser véritablement quant à leur mise en œuvre.

Les modalités de la réglementation de l'IA au niveau international sont controversées. Alors que l'UE et certains de ses États membres plaident en faveur d'une réglementation juridiquement contraignante – même après l'adoption du règlement général sur la protection des données –, d'autres pointent le risque qu'une réglementation trop importante ou trop stricte pourrait faire peser sur le potentiel d'innovation de l'IA ainsi que sur la liberté d'expression et des médias. Quelques organisations internationales (p. ex. le Conseil de l'Europe, l'OCDE et l'UE) ont élaboré de premières normes (juridiquement non contraignantes) relatives à l'IA (voir détails dans le rapport détaillé du groupe de projet). À ce jour, on ignore encore si la communauté internationale parviendra dans un avenir plus ou

[62] Pour un exposé détaillé, voir le rapport du groupe de projet *Internationale Gremien und künstliche Intelligenz*, août 2019. Disponible à l'adresse www.sbfi.admin.ch/ai-f

moins proche à un consensus concernant des normes contraignantes dépassant les seuls principes de base, tout comme on ne sait pas quels seront les processus et les institutions qui joueront un rôle déterminant à plus long terme dans la gouvernance de l'IA.

6.1.3 Activités existantes

Pour un petit pays hautement développé et ouvert sur le monde comme la Suisse, il est essentiel de participer activement au débat sur la gouvernance mondiale de l'IA. C'est pourquoi la Suisse s'implique dans les instances et processus concernés. Il s'agit d'une part des organisations établies telles que l'ONU, l'OCDE, l'UIT, l'UNESCO, l'UE et le Conseil de l'Europe et, d'autre part, de nouveaux groupes et *think tanks* plus modestes qui proposent des travaux de qualité. Ainsi, la Suisse s'est fortement engagée en faveur du lancement du Groupe de haut niveau sur la coopération numérique du Secrétaire général de l'ONU, qui aborde également la question de l'IA. Elle en a soutenu les travaux et considérablement influencé ses résultats[63]. Sur le plan du contenu, la Suisse apporte dans le débat ses valeurs libérales, constitutionnelles et démocratiques ainsi que son expertise. En matière d'IA, elle s'engage tout particulièrement à faire respecter les valeurs et normes fondamentales et établies, telles que les droits de l'homme, et à faire en sorte que toutes les parties prenantes soient impliquées. Dans le même temps, la Suisse prône une approche réglementaire équilibrée, qui permette aussi l'innovation, et œuvre afin que la diversité des approches politiques des États soit prise en compte autant que possible. Dans le cadre de sa politique de la recherche et de l'innovation, la Suisse mise largement sur son système participatif éprouvé, qui vise globalement à créer des conditions-cadre favorables offrant aux acteurs de la recherche et de l'économie la marge de manœuvre nécessaire pour développer leurs propres solutions et axes prioritaires. Avec la Genève internationale, la Suisse bénéficie d'un site qui remplit de nombreuses conditions lui permettant de devenir un pôle de la gouvernance mondiale de l'IA. Genève est néanmoins en concurrence avec d'autres métropoles, dont certaines disposent de ressources beaucoup plus importantes.

6.1.4 Évaluation et actions requises

L'engagement et la position de principe de la Suisse en matière d'IA au sein des instances internationales doivent être poursuivis. Par ailleurs, la Suisse pourrait à l'avenir s'engager encore plus fortement dans certains domaines et Genève s'imposer comme un pôle de la gouvernance mondiale de l'IA.

[63] Une vue d'ensemble complète des représentations de la Suisse dans les instances internationales actives dans le domaine de l'IA est disponible dans le texte et le tableau du rapport complémentaire.

Champ d'action 1 : Échange d'informations et de connaissances et coordination des positions de la Confédération dans les instances internationales Le développement de l'IA intervient dans un environnement mondialisé. Dans ce contexte, les questions de gouvernance se posent de manière variable selon les secteurs, et les interdépendances entre les domaines politiques autrefois distincts ne cessent de se renforcer. Ces développements ne peuvent être encadrés que de manière limitée à l'échelle nationale et requièrent un renforcement de la mise en réseau interdisciplinaire ainsi qu'un échange d'informations et de connaissances au niveau national comme international.	
Utilisation de la « plateforme tripartite » comme réseau de compétences interdisciplinaire national sur les thématiques liées à l'IA dans le domaine de l'IA	La « plateforme tripartite » créée par l'OFCOM en vue de la préparation du Sommet mondial de l'ONU sur la société de l'information (SMSI) doit dès à présent être mobilisée afin d'encourager le dialogue et l'échange d'informations et de connaissances sur les aspects politique, social, économique ou autre de l'IA. Ouverte à toutes les organisations et les personnes intéressées, elle est dotée d'un comité administratif composé de représentants de l'administration fédérale, qui peut être sollicité au besoin pour coordonner les positions de la Confédération dans les instances internationales. Elle peut servir de réseau de compétences interdisciplinaire national sur les questions liées à l'IA, pouvant également opérer une mise en réseau horizontale des connaissances et des expériences et, ainsi, développer des positions cohérentes de la Suisse à l'international. Le tableau des représentations suisses au sein des instances actives dans le domaine de l'IA est régulièrement mis à jour. **Responsabilité :** OFCOM **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 2 : Gouvernance mondiale Il existe des lacunes dans le système de gouvernance du numérique mondial et dans le domaine de l'IA.	
1) Renforcement de la gouvernance mondiale	La gouvernance mondiale doit être renforcée. Pour ce faire, la Suisse doit soutenir et encourager activement le développement de nouveaux modèles, processus et structures de gouvernance et prendre part à la mise en œuvre des recommandations du Groupe de haut niveau sur la coopération numérique du Secrétaire général de l'ONU. **Responsabilité :** OFCOM, DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Intégration de l'IA dans la stratégie de politique étrangère 2020-2023	Les nouvelles technologies comme l'IA ont aussi des implications en matière de politique étrangère. Le renforcement de l'engagement en faveur de la gouvernance mondiale de l'IA doit être examiné dans le cadre de l'élaboration de la stratégie de politique étrangère 2020-2023. **Responsabilité :** DFAE, OFCOM **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 3 : Genève internationale Genève remplit de nombreuses conditions lui permettant de devenir un pôle de la gouvernance mondiale de l'IA. Elle est toutefois en concurrence avec d'autres métropoles, dont certaines disposent de ressources plus importantes.	
1) Amélioration de la mise en réseau et de la coordination des acteurs concernés par l'IA	La Suisse intensifie ses efforts en faveur d'un encouragement à l'amélioration de la mise en réseau et de la coordination des acteurs concernés par l'IA et de la création ou du développement de structures de gouvernance mondiales pour l'IA à Genève. **Responsabilité :** OFCOM, DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Examen du renforcement de la collaboration en vue du « AI for Good Summit »	L'OFCOM, le DFAE et les autorités genevoises discutent avec l'UIT de la possibilité d'une collaboration concernant le développement stratégique du « AI for Good Summit ». **Responsabilité :** OFCOM, DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
3) Renforcement de la Geneva Internet Platform	Le potentiel de la Geneva Internet Platform (GIP) est encore mieux exploité, et la plateforme est renforcée en tant qu'instrument au service de la Suisse. **Responsabilité :** DFAE, OFCOM **Statut :** mise en œuvre dans le cadre des compétences existantes
4) Renforcement de la Genève internationale comme pôle de la gouvernance numérique dans la stratégie de politique étrangère 2020-2023	L'accroissement de l'engagement en faveur du renforcement de la Genève internationale comme pôle de la gouvernance numérique, incluant l'IA, doit être examiné dans le cadre de l'élaboration de la stratégie de politique étrangère 2020-2023. **Responsabilité :** DFAE, OFCOM **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.2 Programme pour une Europe numérique

6.2.1 Vue d'ensemble

Le Programme pour une Europe numérique (DEP) est un nouvel instrument d'encouragement de l'UE. Il sera officiellement lancé en 2021 et s'achèvera fin 2027. Le DEP s'appuie sur cinq piliers : **i)** Calcul à haute performance **ii)** Intelligence artificielle **iii)** Cybersécurité et confiance **iv)** Compétences numériques avancées **v)** Déploiement, meilleure utilisation de la capacité numérique et interopérabilité.

Le présent rapport traitant uniquement de l'intelligence artificielle, nous aborderons ici le deuxième pilier du Programme pour une Europe numérique (DEP). L'objectif du deuxième pilier du DEP est de développer les capacités fondamentales de l'IA, y compris les bases de données et les référentiels d'algorithmes, de les rendre accessibles à toutes les entreprises et administrations publiques et de mettre en réseau les installations d'essai et d'expérimentation existant en matière d'IA dans les États membres de l'UE. Le développement de l'IA requiert donc de grandes quantités de données de qualité. La CE propose que l'UE investisse dans les prochaines années, avec les États membres et le secteur privé, jusqu'à un milliard d'euros dans la création d'un espace européen commun de données, qui mettra des données à la disposition des innovateurs, des entreprises et du secteur public[64]. Afin de promouvoir la collaboration entre le monde scientifique et l'industrie en Europe et d'élaborer un agenda de recherche stratégique commun dans le domaine de l'IA, un nouveau partenariat public-privé en matière de recherche et d'innovation doit en outre être mis en place à long terme. L'expérimentation et les essais en conditions réelles constituent une autre étape importante de la commercialisation des technologies de pointe. Concrètement, cela signifie qu'il faudra investir, dans le cadre du DEP, environ 1,5 milliard d'euros sur l'ensemble du territoire européen en faveur de la création d'installations d'essai et d'expérimentation des produits et services basés sur l'IA de niveau mondial.

6.2.2 Défis

La Suisse examine l'éventualité d'une participation à l'ensemble ou à certains volets du Programme pour une Europe numérique et de son intégration dans le paysage national de la recherche et de l'innovation. Il est à noter qu'il existe des complémentarités et des synergies entre le DEP et plusieurs autres instruments proposés par l'UE pour la période 2021-2027, notamment avec le programme-cadre de recherche « Horizon Europe » de l'UE. La Suisse participe à des projets du programme-cadre de recherche (PCR) de l'UE depuis 1987 et aux accords bilatéraux I en qualité d'État associé depuis 2004. La poursuite de cette participation est prévue à partir de 2021 après l'expiration de la génération actuelle (8e PCR ou « Horizon 2020 ») fin 2020. Du fait de sa liaison étroite avec le programme-cadre de l'UE pour la recherche et l'innovation « Horizon Europe » susmentionnée, la possibilité d'une participation au Programme pour une Europe numérique est traitée dans le message relatif à la participation de la Suisse au paquet Horizon Europe.

Les nouvelles technologies TIC doivent continuer d'être *étudiées et développées* dans le cadre du PCR. Cependant, la *recherche à des fins commerciales, l'innovation et la mise en œuvre sous la forme de produits commercialisables* de même que la diffusion et l'acceptation d'infrastructures et de capacités numériques stratégiques dans les domaines d'intérêt public et dans le secteur privé devront à l'avenir se faire dans le cadre du Programme pour une Europe numérique. C'est la raison pour laquelle la Suisse doit examiner la possibilité de développer les travaux de recherche et d'innovation réalisés dans le cadre du paquet Horizon Europe et les connaissances acquises par ce biais en produits commercialisables en collaboration avec les États de l'UE. Cela se ferait par le biais d'une participation au Programme pour une Europe numérique, qui est ouvert aux pays non-membres de l'UE comme la Suisse.

Les autres défis résident principalement dans le fait que les négociations avec les institutions compétentes de l'UE sont en cours concernant l'élaboration du Programme pour une Europe numérique, les conditions de participation et le budget. Elles ont pris du retard en raison des élections

[64] Commission européenne.« L'intelligence artificielle pour l'Europe », 2018. https://ec.europa.eu/transparency/regdoc/rep/1/2018/FR/COM-2018-237-F1-FR-MAIN-PART-1.PDF

du Parlement européen et du Brexit. On ne sait donc pas encore exactement si, comment et sous quelle forme la Suisse pourra participer au DEP. Il en va de même pour Horizon Europe.

6.2.3 Activités existantes

La Suisse et la Commission européenne examinent actuellement la signature de la Déclaration européenne sur l'intelligence artificielle, qui a déjà été signée par 24 pays européens en avril 2018. La déclaration définit les principaux domaines de l'IA dans lesquels les pays signataires souhaitent collaborer en vue de soutenir la compétitivité de l'Europe en matière d'étude et d'application de l'IA et traiter les questions sociales, économiques, éthiques et juridiques afférentes. Les travaux préparatoires au DEP dans le domaine de l'IA ont déjà commencé sur le plan européen. La stratégie européenne en matière d'IA se fonde sur les atouts scientifiques et industriels de l'Europe et préconise une IA éthique, sûre et de pointe, plaçant les citoyens au cœur de son développement. En décembre 2018, la Commission européenne a présenté un plan d'action coordonné sur l'IA, qui prévoit diverses mesures concrètes, dont certaines doivent déjà être mises en œuvre dans le programme-cadre de recherche Horizon 2020. La Suisse a participé à l'élaboration de ce plan d'action. En juin 2018, la Commission européenne avait institué un groupe d'experts de haut niveau sur l'IA chargé de formuler des directives éthiques ainsi que des recommandations politiques et d'investissement dans le domaine de l'IA. Sur la base de ces travaux a été lancée en avril 2019 une phase pilote visant à déterminer si les directives élaborées par consensus étaient applicables dans la pratique.

Intérêt des acteurs suisses

En février 2019, le SEFRI a réalisé, en consultation avec d'autres organismes fédéraux[65] et la ZHAW, un sondage des parties prenantes suisses afin de connaître leurs positions à l'égard du DEP et leur intérêt quant à une éventuelle participation de la Suisse au programme. Au total, 150 institutions ont pris part à cette enquête en ligne. Les résultats complets ont été publiés dans un rapport distinct du SEFRI[66]. Il faut mentionner que ce sondage visait seulement à évaluer l'intérêt de participants potentiels en Suisse. Pour une participation éventuelle de la Suisse au DEP, la prochaine étape consisterait à en définir, dans un processus distinct, les modalités techniques exactes (association pleine et entière, association partielle, État tiers). Une éventuelle participation au DEP dépend du reste aussi du contexte en matière de politique européenne, de l'issue des négociations avec l'UE et de la situation financière de la Suisse.

L'intérêt des parties prenantes suisses pour le deuxième pilier du DEP s'exprime comme suit : parmi les 150 institutions et experts interrogés, 84 ont répondu qu'ils étaient intéressés par une participation éventuelle au pilier « Intelligence artificielle ». Sur ces 84 institutions, 61 relèvent du secteur public et 23 du secteur privé. Cela montre que le sujet revêt une grande importance tant pour le secteur public que pour le secteur privé. Il est également frappant de constater que ce thème intéresse les institutions et organisations indépendamment de leur taille. Au total, plus de 75 %[67] des institutions ou chercheurs intéressés par le pilier II ont confirmé qu'ils participeraient à des activités liées à l'IA. La volonté de participer effectivement aux activités de recherche dans le domaine de l'IA au niveau européen est particulièrement importante dans le secteur public, avec presque 100 % ; le secteur privé, lui, est plus réservé. Seulement 66 % des sondés pensent qu'un financement fédéral direct (« participation sur le mode projet par projet ») par le SEFRI ou une autre instance à la place d'un encouragement par l'UE est souhaitable. Il est en outre intéressant de relever que les institutions interrogées seraient prêtes à prendre en charge environ 35 % des coûts des projets en moyenne.

6.2.4 Évaluation et actions requises

Comme une initiative de grande ampleur en matière d'IA est déjà en cours à l'échelle européenne dans le cadre du programme Horizon 2020, il convient d'éviter autant que possible les doublons sur le

[65] SECO, OFCOM, armasuisse et la DAE.

[66] SBFI (2019): «Résultats du sondage sur le programme pour une Europe numérique», https://www.sbfi.admin.ch/dam/sbfi/fr/dokumente/2019/07/ergebnisse-dep.pdf.download.pdf/bericht_dep_f.pdf

[67] Deux valeurs ont été calculées dans chaque rapport d'enquête. Une valeur non pondérée qui tient compte du nombre de réponses et une valeur pondérée qui comprend le nombre de personnes couvertes par les réponses.

plan national. Aussi les parties prenantes interrogées se déclarent-elles par exemple en faveur de la participation de la Suisse à la plupart des champs d’action proposés par le plan d’action coordonné européen sur l’IA. Bien que les conditions de participation de la Suisse ne soient pas complètement définies, les parties prenantes tablent sur sa participation au programme Horizon Europe ; au vu des réponses au sondage, la participation au deuxième pilier du DEP semble souhaitée. Les initiatives nationales de recherche dans le domaine de l’IA doivent par conséquent être complémentaires aux initiatives européennes, notamment pour permettre une utilisation efficace des ressources financières et humaines et avec l’objectif d’un « output » maximal. En sa qualité d’office responsable, le SEFRI, en collaboration avec la DAE et l’AFF, en tiendra compte comme il se doit dans la préparation des prochaines étapes.

Champ d’action 1 : Participation de la Suisse à « Horizon Europe » et au « Programme pour une Europe numérique » L’IA est l’un des thèmes centraux du futur programme-cadre pour la recherche et l’innovation « Horizon Europe » et du « Programme pour une Europe numérique ».	
Examen d’une participation à « Horizon Europe » et au « Programme pour une Europe numérique »	Le SEFRI examine, d’entente avec les offices fédéraux concernés (notamment la DAE), si et dans quels domaines une coopération avec l’UE et une participation à ses activités liées à l’IA est souhaitable. **Responsabilité :** SEFRI entre autres **Statut :** mise en œuvre dans le cadre des compétences et des messages existants
Actions supplémentaires requises : non	

6.3 Changements dans le monde du travail

6.3.1 Vue d'ensemble

Comme nous l'avons vu au chapitre 2.3, l'intelligence artificielle (IA) est réputée posséder le potentiel d'une technologie de base qui, comme la machine à vapeur ou l'électricité, peut pénétrer et bouleverser des branches ou des économies entières. Il est incontestable que cette évolution technologique donne des impulsions importantes pour le développement économique. Dans le même temps, l'IA est vue comme une source de nouveaux risques. Les répercussions négatives possibles de l'IA sur l'emploi reviennent de façon récurrente dans les discussions. On entend régulièrement s'exprimer la crainte de voir l'IA un jour en mesure de remplacer l'intelligence humaine et de rendre ainsi pratiquement toutes les activités automatisables. Quels risques et opportunités les développements dans le domaine de l'IA représentent-ils pour le marché suisse du travail ? Quelles sont les facteurs décisifs permettant une adaptation réussie au changement technologique ?

6.3.2 Défis

Concernant le marché du travail, le développement et la propagation de l'IA ne sont pas isolés des autres moteurs de l'évolution technologique. À l'heure actuelle, aucun élément ne semble par exemple suggérer que l'IA transformera fondamentalement le marché du travail d'une manière différente que les développements technologiques précédents, en particulier les technologies de la numérisation.

L'utilisation des novelles technologies influence l'évolutation globale de l'emploi au travers de différents canaux. Le progrès technique en général et le développement des technologies numériques en particulier contribuent à améliorer la productivité et à réduire les coûts. A chaque fois que cela se justifie d'un point de vue économique, les étapes de travail concernées seront automatisées à moyen ou long terme. Si ces processus de substitution vont trop vite ou que les profils de compétences proposés ne s'adaptent pas à la demande de travail ou le font trop lentement, ils peuvent engendrer une inadéquation des qualifications et du chômage technologique au niveau global.

Outre les effets de substitution, les technologies d'automatisation peuvent aussi être complémentaires aux emplois actuels. Ces technologies complémentaires revalorisent le portefeuille d'activités de la main-d'œuvre et accroissent la productivité du travail, ce qui peut favoriser des hausses de salaire. Avec l'IA, davantage de professions exigeantes en termes de faculté de jugement et de pensée critique pourraient profiter de cet effet complémentaire.

De plus, le progrès technologique représente un facteur de stimulation de la demande globale. L'amélioration de la productivité et la diminution des coûts de production entraînent généralement une baisse des prix des produits qui se répercute favorablement sur le revenu disponible réel des consommateurs. L'augmentation de la demande stimule la production et accroît les besoins de main-d'œuvre. Il est toutefois difficile de quantifier cet effet sur la demande globale, puisque l'ampleur du phénomène dépend fortement de facteurs comme la sensibilité de la demande aux prix ou la propension à consommer des ménages. Enfin, le développement de nouvelles technologies et applications a un effet direct positif sur l'emploi.

Selon les estimations actuelles, 14 % des emplois sont considérés comme étant fortement menacés par le processus d'automatisation dans les pays de l'OCDE[68]. L'OCDE estime que la probabilité d'une évolution négative de l'emploi à un niveau agrégé est faible. Pour l'OCDE, le principal défi est que la structure d'activité de 32 % supplémentaires des emplois peut être modifiée de manière significative dans les prochaines décennies.

6.3.3 Activités existantes

Le Conseil fédéral a récemment examiné de manière approfondie les répercussions de la numérisation sur l'économie et le marché de l'emploi en Suisse[69]. La Suisse bénéficie globalement d'un contexte

[68] Cf. OCDE. *The Future of Work – Employment Outlook 2019*, 2019. Disponible à l'adresse https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2019_9ee00155-en

[69] Citons les rapports *Principales conditions-cadre pour l'économie numérique* du 11 janvier 2017 et *Conséquences de la numérisation sur l'emploi et les conditions de travail* du 8 novembre 2017. Le Conseil fédéral a également publié un rapport sur les défis de la numérisation pour la formation et la recherche, ainsi

favorable pour tirer également parti de l'IA. Selon le Conseil fédéral, il est important de préserver l'attrait de la place économique suisse par la stabilité des conditions économiques générales, une politique monétaire axée sur la stabilité et une réglementation flexible du marché de l'emploi, alliée à un partenariat social bien rodé et une politique active du marché du travail. Par ailleurs, il est essentiel de maintenir la grande capacité d'innovation des entreprises suisses en comparaison internationale, qui est notamment favorisée par la qualité et le degré de perméabilité élevés du système suisse de formation.

Les développements futurs dans le domaine doivent être suivis de près. C'est pourquoi le Conseil fédéral a mis en place un monitorage des conséquences de la transformation numérique sur le marché du travail en novembre 2017. Les résultats du monitorage devront faire l'objet d'un rapport après cinq ans (échéance : fin 2022) et permettront ainsi une identification précoce des nouveaux enjeux.

6.3.4 Évaluation et actions requises

Bien que la transformation numérique soit un processus engagé depuis longtemps, aucun recul de l'emploi total imputable aux nouvelles technologies n'a été constaté dans les pays développés. Jusqu'à présent, les emplois supprimés ont toujours au minimum été compensés dans d'autres domaines. Dans son rapport consacré aux conséquences de la numérisation sur l'emploi et les conditions de travail, le Conseil fédéral indique que l'on peut s'attendre à ce que le changement structurel lié à la numérisation crée de nouvelles opportunités de travail et favorise ainsi une augmentation globale de l'emploi. Au vu des changements observés au cours des vingt dernières années, on peut toutefois penser que le monde du travail évoluera de manière significative dans les décennies à venir, que ce soit au niveau des branches, des métiers ou des activités. Dans le même rapport, le Conseil fédéral constate en outre que, malgré les progrès sensibles réalisés dans le domaine de l'IA, aucun changement structurel dont le rythme serait supérieur à la moyenne ne s'observe actuellement au niveau global. Dans les prochaines années, il est plus probable que l'on assiste à un développement progressif dans le cadre du changement structurel en cours qu'à une révolution technique disruptive.

L'économie suisse doit tirer parti du potentiel des nouvelles technologies pour accroître la productivité et la croissance. L'IA prendra de plus en plus d'importance. L'évolution technologique offre à la place économique suisse, qui est orientée vers l'innovation et les produits à forte création de valeur la chance, de gagner encore en compétitivité et de préserver ainsi pour les travailleurs des possibilités d'emploi intéressantes, voire d'en créer de nouvelles. Pour ce faire, il est crucial de sauvegarder les facteurs de réussite permettant de faire face au changement structurel.

Les qualifications et compétences des travailleurs doivent être ajustées suffisamment tôt aux nouveaux besoins du marché du travail afin d'éviter qu'une inadéquation ne s'instaure en la matière à la suite de l'évolution technologique. Une orientation rapide de la formation et une adaptation des filières de formation à l'évolution des exigences sont indispensables. L'apprentissage tout au long de la vie sous la forme d'une formation et d'un perfectionnement permanents jouera un rôle de plus en plus important. Le Conseil fédéral est conscient de ce défi et a déjà engagé des mesures pour y répondre, par exemple la promotion des compétences de base sur le lieu de travail (cf. chapitre 6.5).

Dans ce contexte, s'agissant des conséquences de l'intelligence artificielle sur le marché du travail, aucune nouvelle mesure ne s'impose.

Champ d'action 1 : Conséquences de l'IA sur le marché du travail	
Suivi des développements sur le marché du travail	Le SECO observera les enjeux dans ce contexte et traitera les questions soulevées dans le cadre des compétences existantes. **Responsabilité :** SECO **Statut :** suivi dans le cadre des compétences existantes
Actions supplémentaires requises : non	

qu'un autre rapport sur les conséquences de la numérisation sur la fiscalité et le financement des assurances sociales. Par ailleurs, il a mené une vaste enquête, baptisée « Test de compatibilité numérique », visant à identifier les obstacles que la réglementation actuelle en matière de politique économique pose inutilement à la numérisation.

6.4 L'intelligence artificielle dans l'industrie et les services[70]

6.4.1 Vue d'ensemble

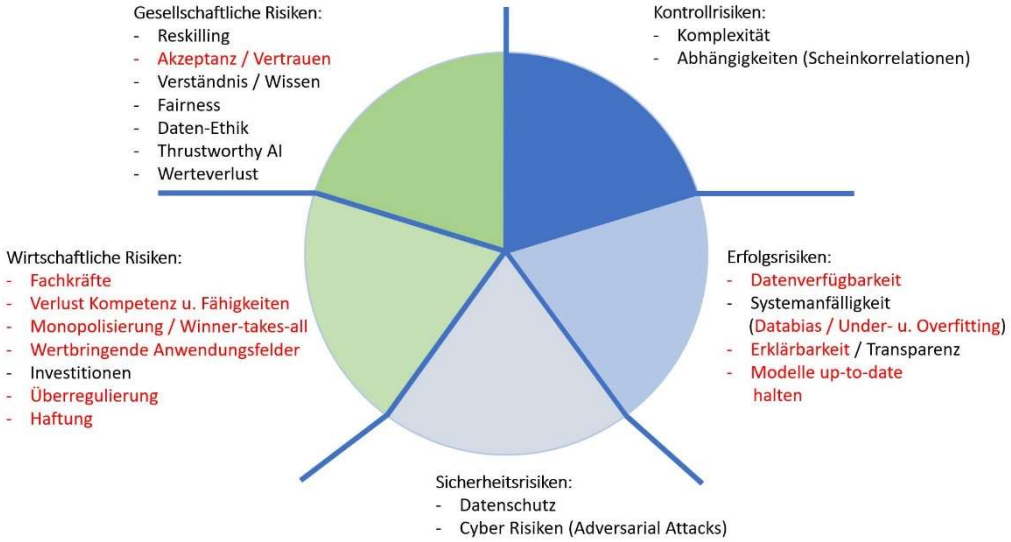
Dans le contexte de la quatrième révolution industrielle, l'IA offre de nouvelles possibilités de tirer une valeur ajoutée du potentiel représenté par la mise en réseau numérique et du flux de données qui en découle. Sur le plan coûts, cela permet aux entreprises et aux prestataires de services qui utilisent des applications IA de réduire leurs charges internes. Outre l'utilisation de l'IA dans les domaines dans lesquels les coûts peuvent être réduits, une expérience client qui serait améliorée par l'IA recèle un potentiel supplémentaire pour les entreprises. Au niveau des recettes, les applications IA permettent aux entreprises et aux prestataires de services de proposer des produits et services nouveaux ou améliorés.

L'utilisation de l'IA dans l'industrie 4.0 et le secteur des services n'en est qu'à ses débuts. À l'heure actuelle, de nombreux projets menés en Suisse le sont à titre expérimental, et les entreprises investissent dans l'IA pour accumuler de l'expérience. L'IA exécute des tâches liées à la perception qui jusqu'à présent ne pouvaient être exécutées que par l'homme (p. ex. tâches répétitives), des tâches d'optimisation ou d'automatisation des processus. Avec l'IA, les processus sont plus efficaces (en termes de temps et de coûts) et évolutifs.

6.4.2 Défis

De l'avis des spécialistes consultés, les défis auxquels sont confrontées les entreprises peuvent être classés, d'une part, dans les défis techniques (tels que les risques de contrôle ou de sécurité) et, d'autre part, dans les défis sociaux (p. ex. acceptation/confiance) et économiques (notamment personnel qualifié, surréglementation). La figure 13 présente les risques variés supportés par les entreprises. Les défis spécifiques auxquels les entreprises accordent généralement une priorité élevée sont signalés en rouge.

Figure 13 : Défis des systèmes d'IA pour les entreprises



Source : SATW, « Künstliche Intelligenz in Industrie und Dienstleistungen », 2019.

6.4.3 Activités existantes

Le secteur économique s'organise au moyen de nombreuses initiatives ayant un lien direct ou indirect avec l'IA (quelques exemples en sont donnés dans le tableau 4). Au-delà de ces activités, les

[70] La section portant sur l'importance et les défis de l'IA dans l'industrie et les services a été rédigée par la SATW sur mandat du SEFRI. À cet effet, des experts issus de la recherche et de l'économie ont été interrogés. Le texte reflète l'opinion des experts. Pour un exposé détaillé, voir SATW. *Künstliche Intelligenz in Industrie und Dienstleistungen*, rapport rédigé sur mandat du Secrétariat d'État à la formation, à la recherche et à l'innovation, 2019. Disponible à l'adresse www.sbfi.admin.ch/ai-f

entreprises mènent leurs propres projets dans le domaine de l'IA[71]. Ces derniers portent entre autres sur des aspects tels que les directives éthiques, l'utilisation responsable de l'IA ou d'autres thématiques similaires visant à accroître la confiance dans les produits et services faisant appel à l'IA.

Tableau 4 : Initiatives et activités en Suisse dans le domaine de l'intelligence artificielle (à titre d'exemple)

Swiss Alliance for Data-Intensive Services	Dans le cadre de cette initiative, le groupe d'experts « Data Ethics » élabore un *Ethical Codex for Data-Based Value Creation* destiné à affermir la confiance dans la technologie.
Swiss Group of Artificial Intelligence and Cognitive Science	Groupe spécialisé de la Société suisse d'informatique qui œuvre en faveur des échanges entre chercheurs, utilisateurs et personnes intéressées dans le domaine de l'IA.
SwissCognitive	Initiative lancée par de nombreuses entreprises de l'industrie et du secteur des services entièrement dédiée à l'IA, servant de plateforme d'échange et de réseau.
Digitalswitzerland	Initiative nationale composée de plus de 150 membres visant à conforter la position de la Suisse en tant que pôle d'innovation de pointe.
Industrie 2025	Initiative sectorielle suisse qui informe, sensibilise, met en réseau et aide les acteurs concernés au sujet de l'industrie 4.0.
Swiss Smart Factory	Plateforme du Switzerland Innovation Park Biel/Bienne pour les questions interdisciplinaires en lien avec l'industrie 4.0.

Source : SATW, « Künstliche Intelligenz in Industrie und Dienstleistungen », 2019.

L'utilisation de l'IA intervient dans le contexte du développement de la numérisation. À travers sa stratégie **« Suisse numérique »**, la Confédération soutient la numérisation de l'ensemble du pays dans tous les secteurs. Par ailleurs, un grand nombre de mesures en faveur du renforcement des compétences numériques dans la formation, la recherche et l'innovation ont été engagées dans le cadre du **plan d'action « Numérisation pour le domaine FRI »**. La Confédération agit déjà contre la pénurie de main-d'œuvre qualifiée dans les métiers essentiels dans le contexte de la numérisation, par exemple avec l'**initiative visant à combattre la pénurie de personnel qualifié** qui s'est achevée en 2018.

6.4.4 Évaluation et actions requises

Selon les experts, les principaux défis liés à l'utilisation de l'IA dans l'industrie et le secteur des services doivent être relevés par les acteurs économiques eux-mêmes, en particulier pour ce qui concerne les enjeux techniques découlant de l'IA. L'économie s'est attaquée à ces défis à travers de nombreuses initiatives, elles-mêmes soutenues par plusieurs initiatives publiques. Parmi elles, on trouve aussi des initiatives visant à établir la confiance dans les technologies IA et à promouvoir une utilisation responsable de ces dernières.

Les experts interrogés dans le cadre de ce rapport identifient quelques défis généraux qui concernent la Confédération en rapport avec l'utilisation de l'IA dans l'industrie et le secteur des services (un compte-rendu exhaustif se trouve dans le rapport SATW, « Künstliche Intelligenz in Industrie und Dienstleistungen », 2019. Néanmoins, la plupart de ces défis ne se rapportent pas exclusivement à l'IA, mais à la numérisation de manière générale. Aussi certains d'entre eux ont-ils déjà été identifiés et traités dans le cadre de la réponse à la numérisation et au processus de transformation numérique.

La suggestion des experts recommandant que la Confédération élabore un **plan stratégique national commun sur l'IA** se traduit, au niveau de la Confédération, par l'élaboration de lignes stratégiques à la suite du présent rapport. De plus, les politiques concernées par l'IA sont traitées comme thème prioritaire dans le cadre de la stratégie « Suisse numérique », des initiatives privées pertinentes pouvant également être prises en considération. Pour lutter contre la pénurie de main-d'œuvre qualifiée, les experts préconisent d'améliorer les conditions-cadre applicables à l'embauche de la

[71] Citons entre autres Microsoft, Google, PwC et D-One.

main-d’œuvre étrangère formée en Suisse. La mise en œuvre actuelle de la **motion Dobler 17.3067**, entre autres, y contribue en allégeant les conditions d’embauche des étudiants originaires de pays tiers formés en Suisse. Pour le moment, la nécessité de créer d’autres champs d’action ne se fait donc pas sentir.

<table>
<tr><td colspan="2">Champ d’action 1 : L’IA dans l’économie
L’IA a le potentiel d’améliorer considérablement l’efficience dans la production et la fourniture de services, et permet une plus forte individualisation afin de proposer aux clients des solutions sur mesure.</td></tr>
<tr><td>Suivi des développements de l’IA dans l’industrie et les services</td><td>Le SECO observera dans ce contexte les enjeux liés à l’utilisation de l’IA dans l’industrie et le secteur des services et traitera les questions soulevées dans le cadre des compétences existantes.

Responsabilité : SECO
Statut : suivi dans le cadre des compétences existantes</td></tr>
<tr><td colspan="2">Actions supplémentaires requises : non</td></tr>
</table>

6.5 L'intelligence artificielle dans la formation[72]

6.5.1 Vue d'ensemble[73]

Impact sur l'enseignement et l'apprentissage grâce à l'intelligence artificielle : L'intelligence artificielle (IA) amène des énormes opportunités d'améliorer l'enseignement et l'apprentissage. Son utilisation dans la pratique a été relativement modeste jusqu'à aujourd'hui, mais la situation pourrait évoluer rapidement.[74] L'IA élargit les possibilités d'automatisation de correction d'exercices ou d'évaluation de tests et permet d'analyser les traces afin de comprendre lecomportement de l'étudiant grâce au « Learning Analytics ». Ceci permet d'affiner l'adaptation de l'apprentissage aux besoins individuels de chaque élève ou de prédire un futur échec. Des informations détaillées concernant les améliorations de l'enseignement et de l'apprentissage sont disponibles au chapitre 1 du rapport complémentaire « L'intelligence artificielle dans la formation ».

6.5.2 Défis

Compétences requises pour l'utilisation et la production de systèmes intelligents : L'intelligence artificielle sera de plus en plus présente dans notre vie, tant privée que professionnelle, ce qui a des conséquences sur les compétences dont les citoyennes et les citoyens devraient disposer pour vivre et travailler dans une société numérisée. Il est difficile de prévoir avec précision quelles seront les compétences nécessaires dans le futur. Il est certain que l'IA créera des nouvelles professions et en fera probablement disparaître d'autres. Les compétences numériques (de base et avancées) seront en outre de plus en plus demandées. En revanche, la numérisation met à risque les professions les moins qualifiées qui comportent des tâches répétitives facilement automatisables. À noter que les « soft skills » ou compétences transversales, telles que la curiosité, la créativité, la collaboration, l'empathie, le leadership ou encore la résolution de problèmes, seront de plus en plus demandées car exclues aujourd'hui du champ de compétences des machines.

Par rapport aux compétences numériques (de base ou avancées) qui sont désormais nécessaires dans presque tous les secteurs à cause de la numérisation, l'intelligence artificielle requiert des compétences renforcées dans certains domaines spécifiques. Par exemple, les métiers qui travaillent avec l'IA ont besoin d'un niveau très élevé de compétences en algorithmique et mathématiques, ainsi qu'un renforcement de la pensée critique et du raisonnement éthique. Pour les métiers qui se limitent à utiliser les systèmes intelligents, il faut une compréhension intuitive des algorithmes afin d'en saisir les possibilités et les limites. Il est important de souligner que l'IA pourrait avoir une influence sur tous les secteurs économiques, y compris sur les métiers manuels ou hautement qualifiés (comme la profession d'avocat). Voir le chapitre 2.1 du rapport complémentaire pour plus de détails sur l'évolution des compétences.

L'égalité des chances et la hausse de la participation des femmes dans les secteurs de la recherche, de la production et des applications IA sont un point auquel il faut rester attentif. Il convient en effet de renforcer la motivation des personnes de sexe féminin pour les formations et les professions du domaine MINT (mathématiques, informatique, sciences naturelles, technique).

Défis pour la pratique pédagogique : L'effet réel des systèmes d'intelligence artificielle demeure restreint aujourd'hui à cause de l'utilisation assez limitée dans la pratique actuelle. Il faut préciser que les données relevant du domaine pédagogique sont des informations sensibles à caractère personnel. Des problèmes se posent plus particulièrement sur la question des lacunes dans l'explicabilité des résultats et sur les biais contenus dans les algorithmes. Cf. chapitre 2.2 du rapport complémentaire « L'intelligence artificielle dans la formation » pour plus d'informations sur les défis pédagogiques.

À noter que le rôle formatif, pédagogique et social de l'école n'est pas du tout remis en question. La figure de l'enseignant ou de l'enseignante ne perdra pas d'importance avec l'arrivée de l'IA, au contraire ; le personnel enseignant pourra se permettre de moins se focaliser sur la correction

[72] Pour un exposé détaillé, voir Secrétariat d'État à la formation, à la recherche et à l'innovation, « L'intelligence artificielle dans la formation », 2019. Disponible à l'adresse www.sbfi.admin.ch/ai-f

[73] Le présent texte se fonde en grande partie sur un article de Pierre Dillenbourg, prof. ordinaire en technologies de formation à l'EPFL, sur mandat du SEFRI.

[74] Tuomi, I. *The Impact of Artificial Intelligence on Learning, Teaching, and Education. Policies for the future*, éd. Cabrera M., Vuorikari R. et Punie Y., Office des publications de l'Union européenne, Luxembourg, 2018.

d'exercices ou d'autres tâches administratives qui pourraient êtreexécutées par des machines pour se consacrer à la préparation des cours ou fournir un soutien ciblé à chaque élève.

Implications éthiques de l'utilisation de l'IA dans l'enseignement : L'utilisation croissante de données soulève également des questions éthiques. À l'avenir, qui devra être autorisé à recueillir et à traiter des données auprès de qui, en quelles quantités et à quelles fins ? Quelle doit être la culture des données dans le domaine de la formation ? Sur quels principes doit-elle reposer ? L'une des principales missions d'un système intégré et sécurisé de collecte et d'analyse de données standardisées dans le secteur éducatif est de renforcer la confiance dans cette utilisation et d'accroître ainsi son acceptation auprès des parties prenantes concernées. Cette confiance est une condition nécessaire pour que les connaissances tirées d'éventuelles analyses de données soient perçues comme utiles et justifiées.[75]

Lacunes dans la recherche en éducation : Afin d'améliorer les systèmes intelligents utilisés dans l'enseignement et l'apprentissage, il est important de poursuivre la recherche en éducation afin de résoudre les problèmes techniques et éthiques cités précédemment. La collaboration entre les experts en IA et les chercheurs en éducation est aussi importante afin de tenir compte des spécificités de l'éducation.

6.5.3 Activités existantes

Afin de déterminer les conséquences de l'IA sur les compétences et le système de formation, il est essentiel de ne pas observer les développements de manière isolée, mais de les appréhender dans le contexte général de la numérisation. Ceci est d'autant plus important vue la difficulté de prévoir comment évolueront les exigences en termes de compétences et quelles aptitudes seront nécessaires dans la pratique pour l'utilisation des outils de l'intelligence artificielle.

À tous les niveaux de la formation les autorités cantonales et la Confédération, ainsi que les établissements de formation, sont conscient des défis que pose la numérisation. L'évolution dans ce domaine est suivie de près pour identifier les risques à temps et exploiter pleinement le potentiel de l'IA. Plusieurs stratégies, initiatives et mesures sont déjà prévues ou en cours de réalisation. Ces stratégies concernent la numérisation de l'éducation en général, mais il est clair qu'elles visent également à favoriser une utilisation responsable de l'IA et à assurer la transmission des compétences adéquates dans ce domaine. Le chapitre 3 du rapport complémentaire montre les stratégies et les mesures mises en œuvre à tous les niveaux de la formation ainsi qu'une liste non exhaustive des activités existantes dans le domaine de l'IA par niveau de la formation.

Au niveau international, force est de constater que le thème des compétences et de l'éducation est omniprésent dans les réflexions concernant l'IA. Par exemple, les recommandations sur l'IA adoptées par l'OCDE en mai 2019 couvrent aussi le domaine des compétences. De son côté, l'un des objectifs du « Coordinated Plan on Artificial Intelligence »[76] adopté par la Commission européenne en décembre 2018 consiste à adapter les programmes et systèmes de formation afin de mieux préparer la société à l'IA**.** Les « Policy and investment recommendations for trustworthy Artificial Intelligence » publiées le 26 juin 2019 par l'UE se réfèrent aussi aux compétences. Enfin, la promotion des compétences numériques dans des domaines d'avenir tels que les IA est l'un des volets essentiels du « Programme pour une Europe numérique » (DEP), qui sera mis en œuvre de 2021 à 2027 (cf. chapitre 6.2 sur le DEP). Le SEFRI participe à différents groupes de travail de la Commission européenne, comme le « Working Group Digital Education » ou le « Working Group VET», qui s'occupent aussi d'intelligence artificielle. La Suisse est ainsi bien informée sur les activités menées par les voisins européens dans le domaine de l'intelligence artificielle dans l'éducation. L'UE et ses États membres ont eux aussi reconnu l'importance de cette technologie et discutent actuellement sur les potentiels et les défis. Cependant, pour le moment, seulement certains pays (p. ex. Finlande) ont lancé des stratégies ou des initiatives spécifiques au domaine de l'IA dans la formation.

[75] educa.ch. « Données dans l'éducation – données pour l'éducation. Bases et pistes de réflexion d'une politique d'utilisation de données pour l'espace suisse de formation ». Berne. 2019.
[76] Communication « Un plan coordonné dans le domaine de l'intelligence artificielle », COM(2018) 795.

6.5.4 Évaluation et actions requises

Dans le respect des compétences respectives, Confédération et cantons collaborent aujourd'hui étroitement dans le cadre du comité de coordination « Numérisation de l'éducation » afin d'assurer de bonnes conditions-cadre dans le domaine de la numérisation de l'éducation. Dans le domaine des hautes écoles, la Confédération et les cantons collaborent au sein des organes de la Conférence suisse des hautes écoles (CSHE). Les défis de la numérisation pour le secteur de la formation ont été traités par les instances compétentes, qui ont déjà engagé un grand nombre de mesures en ce sens. Ces travaux englobent également la question de l'IA. Dans le cas spécifique des hautes écoles, s'appuyant sur la planification stratégique de swissuniversities, la Conférence suisse des hautes écoles (CSHE) a fixé en mai 2019 un ordre de priorités politiques à l'échelle nationale pour la période 2021-2024. Ces priorités englobent l'encouragement de la relève dans les disciplines du domaine MINT, et plus spécifiquement de la formation de personnes qualifiées en TIC et du renforcement de compétences numériques chez les jeunes diplômés et le personnel scientifique. À l'heure actuelle, nous n'identifions aucun besoin de clarifications ou d'instances supplémentaires en plus de ces mesures. À noter que la numérisation est l'un des domaines prioritaires du programme d'action « Vision 2030 » et sera un domaine transversal de grande importance dans le message FRI 2021-2024. Des réflexions sur le thème spécifique de l'intelligence artificielle se tiendront aussi dans ce contexte. Dans le cadre des structures existantes et du dialogue avec les acteurs, les thématiques suivantes doivent être traitées ou approfondies selon les besoins.

<table>
<tr><td colspan="2">Champ d'action 1 : Assurer la transmission des compétences adéquates
Les compétences nécessaires à l'utilisation de l'IA doivent être acquises pendant le parcours scolaire ainsi que par la formation continue tout au long de la vie afin d'éviter une polarisation de la société entre les gens disposant de ces compétences et les autres. La compréhension générale des algorithmes est fondamentale, mais le sont également les « soft skills » (ou compétences transversales).
Le système de formation doit également assurer l'acquisition adéquate de compétences spécifiques à la production de l'IA et, partant, contribuer à la formation de spécialistes de l'IA.
Il faut renforcer la motivation et la formation des personnes de sexe féminin pour les disciplines du domaine MINT.</td></tr>
<tr><td>Assurer la transmission des compétences nécessaires à l'utilisation de l'intelligence artificielle à tous les niveaux de la formation</td><td>Le SEFRI assure une transmission des compétences nécessaires à l'utilisation de l'intelligence artificielle aux niveaux de la formation pour lesquels il est responsable (p. ex. formation professionnelle). Pour les autres niveaux de la formation, le SEFRI collabore étroitement avec les cantons au sein des organes où il est représenté.
Responsabilité : SEFRI, cantons et autres acteurs pertinents du domaine
Statut : mise en œuvre dans le cadre des compétences existantes</td></tr>
<tr><td colspan="2">Actions supplémentaires requises : non</td></tr>
</table>

Champ d’action 2 : **Assurer une utilisation transparente et responsable de l’IA dans la formation** L’utilisation de l’IA pour l’enseignement et l’apprentissage soulève plusieurs questions règlementaires relatives à l’accès, à la collecte et à l’utilisation des données générées pendant le processus d’apprentissage. Afin de pouvoir profiter des bienfaits de l’IA, il faut assurer une utilisation transparente et responsable de l’IA dans la formation. En vue d’atteindre ce but, il faut réduire les risques concernant la sécurité et la protection des données, la protection de la sphère privée et les aspects éthiques.	
Assurer une utilisation transparente et responsable de l’IA dans la formation	Le SEFRI collabore étroitement avec les cantons au sein des organes où il est représenté pour l’analyse d’éventuelles mesures nécessaires à garantir une utilisation transparente et responsable de l’IA dans la formation. **Responsabilité :** SEFRI, cantons et autres acteurs pertinents du domaine **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.6 L'intelligence artificielle dans la science et la recherche[77]

6.6.1 Vue d'ensemble

Dans le domaine de la recherche, l'IA est applicable à large échelle dans toutes les disciplines, permet une nouvelle lecture des connaissances existantes et recèle un potentiel considérable de nouvelles découvertes. Par exemple, elle permet l'étude en direct de processus auxquels l'homme ne pourrait pas accéder sans elle. En science, l'IA est utilisée pour générer et analyser des données ou améliorer des méthodes. Par ailleurs, les programmes de traduction basés sur l'IA rendent les publications en langue étrangère accessibles à un public plus large, et les formes de revue de la littérature basées sur l'IA requièrent des délais de traitement plus courts. En conséquence, le travail scientifique est plus efficace et plus productif : le nombre d'essais nécessaires diminue, et les processus peuvent être modélisés plus rapidement.

6.6.2 Défis

Selon les experts, le recours accru à l'IA par la science et la recherche s'accompagne de nombreux défis. Il existe, d'une part, des défis techniques et, d'autre part, des défis soulevant des questions générales d'ordre juridique et social.

Tableau 5 : Défis liés à l'IA spécifiques à la science et à la recherche

Explicabilité	L'*Explainable AI* ou « IA explicable » ne cesse de gagner en importance dans la science et la recherche.
Confiance dans les systèmes d'IA	Le degré d'autonomie des systèmes d'IA doit être défini selon le domaine d'application.
Taux d'erreur	Les distorsions (« biais ») sont difficiles à identifier.
Formation	La formation est essentielle pour comprendre les systèmes d'IA et interpréter correctement les résultats.
Acteurs privés	Les acteurs privés sont importants pour la recherche, mais il peut en résulter une dépendance aux outils logiciels.
Protection des données	Il est nécessaire d'adopter des directives de protection des données et de la sphère privée les mieux adaptées possible.
Collecte des données	Des directives définissant les données et les outils IA (en ligne) dont l'utilisation est autorisée doivent être publiées.
Propriété intellectuelle	Il convient de clarifier les questions relatives à la propriété intellectuelle ainsi qu'à la protection et à la disponibilité des données.
Accessibilité	Le secteur de l'édition doit être adapté aux besoins spécifiques de l'IA (p. ex. accès aux publications et *text mining*).
Infrastructures	Les infrastructures nécessaires pour le développement de compétences dans le domaine de l'IA doivent pouvoir être financées.

Source : SATW, « Künstliche Intelligenz in Wissenschaft und Forschung », 2019.

[77] Cette section consacrée à l'IA dans la science et la recherche a été rédigée par la SATW sur mandat du SEFRI. La SATW a interrogé sept experts issus de différentes disciplines dans le cadre d'entretiens structurés. Le rapport rédigé sur cette base a ensuite été étudié et consolidé en concertation avec les spécialistes de l'IA proposés par swissuniversities. Cette section a été finalisée par le SEFRI. Le texte reflète l'opinion des experts. Pour un exposé détaillé, voir SATW. *Künstliche Intelligenz in Wissenschaft und Forschung*, rapport rédigé sur mandat du Secrétariat d'État à la formation, à la recherche et à l'innovation, 2019. Disponible à l'adresse www.sbfi.admin.ch/ai-f

6.6.3 Activités existantes

Il existe en Suisse un grand nombre d'institutions et d'initiatives qui s'intéressent de près à la recherche sur l'IA et ont largement contribué au développement des technologies IA au cours des décennies passées. Au plan international, les compétences et les connaissances dans le domaine de l'IA sont développées, les talents encouragés et le transfert de connaissances intensifié. Les acteurs suisses sont aussi fortement impliqués dans ces initiatives et y jouent dans certains cas un rôle de premier plan. Le chapitre 5 donne quelques exemples d'activités existantes.

L'utilisation de l'IA dans le traitement des données et publications de recherche nécessite que celles-ci soient disponibles facilement et le plus librement possible. À cet effet, les hautes écoles suisses ont mis en place des mesures dans les domaines *Open Access to Publication* et *Open Research Data*. En collaboration avec le FNS, elles ont élaboré une **stratégie nationale sur l'Open Access** visant à ce que toutes les publications scientifiques financées par des fonds publics soient librement accessibles d'ici 2024 (dès 2020 pour les publications résultant de projets financés par le FNS). Au niveau européen, la **plateforme BEAT** est une infrastructure informatique européenne dédiée à l'Open Science.

6.6.4 Évaluation et actions requises

À l'heure actuelle, les spécialistes de la Confédération n'identifient aucun besoin de mesures ni d'instances supplémentaires. Dans le cadre des structures existantes et du dialogue avec les acteurs, les thématiques suivantes doivent être traitées ou approfondies selon les besoins[78].

Développement et applications des méthodes IA : les défis de l'IA doivent être traités directement par le monde scientifique et les hautes écoles, qui s'y attellent déjà de manière intensive. En particulier, il convient de renforcer l'interdisciplinarité et les échanges concernant les méthodes IA traditionnelles et « nouvelles ». Dans ce cadre, les membres des hautes écoles peuvent faire office « d'infopoints en matière d'IA», en interne comme en externe. Ce sont les hautes écoles qui définissent elles-mêmes les axes de recherche et les nouvelles initiatives (par exemple, la « Digital Society Initiative » de l'Université de Zurich). La Confédération ne peut donner aucune directive en la matière.

Questions générales relevant de la compétence du domaine des hautes écoles : les questions générales doivent être clarifiées par les instances existantes de coordination (politiques et universitaires) du domaine des hautes écoles, notamment : (i) dialogue concernant les défis et les recommandations dans le domaine de l'IA ; (ii) examen de la nécessité et, le cas échéant, élaboration d'une stratégie fixant les règles d'utilisation de l'IA ; clarification du rôle des acteurs privés et des questions liées aux données (*owner-/stewardship*) ; (iii) thématiques déjà traitées par les hautes écoles : accessibilité (*Open Science / Open Access*) et infrastructures (référentiels) ; (iv) intervention de commissions d'éthique sur les questions générales.

Développement accru des compétences dans le domaine de l'IA : la Suisse dispose déjà d'instruments de qualité en matière de promotion de l'IA. Un programme d'encouragement isolé serait peu pertinent, car de nombreuses disciplines seraient concernées, et l'encouragement ne pourrait pas être ciblé. Les compétences numériques générales doivent être renforcées dans le cadre des structures et des organes existants ; ce renforcement a déjà été abordé au niveau de la Confédération dans le plan d'action « Numérisation pour le domaine FRI » et à travers les instruments ouverts et compétitifs de la Confédération. La numérisation et l'IA ont été définies comme des axes centraux de la planification stratégique des hautes écoles pour la période 2021-2024.

[78] Les recommandations des experts concernent également les domaines des données, du droit et de la propriété intellectuelle ; pour les développements de ces aspects, il est envoyé aux sections respectives du présent rapport.

Champ d’action 1 : Compétences dans la recherche et l’innovation La recherche et la formation jouent un rôle crucial dans la maîtrise des défis de l’IA. Le maintien des compétences au plus haut niveau doit donc être garanti.	
Préservation des compétences en recherche et TST dans le cadre de la politique FRI	Les mesures du plan d’action « Numérisation pour le domaine FRI » sont intégrées à la période FRI 2021–2024 et mises en œuvre par les acteurs de manière autonome. Dans le cadre des structures et organes existants dédiés à la numérisation, la politique FRI doit en outre garantir que les acteurs du monde scientifique et du TST sont préparés aux défis de l’intelligence artificielle et les relèvent au moyen de leurs activités et de leurs compétences respectives en matière de numérisation. **Responsabilité :** SEFRI **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.7 L'intelligence artificielle dans la cybersécurité et la politique de sécurité[79]

6.7.1 Vue d'ensemble

Du point de vue de la politique de sécurité, il y a trois domaines thématiques où l'IA joue un rôle de plus en plus important : (i) politique de sécurité extérieure et gouvernance internationale ; (ii) forces armées et évolution de la conduite de la guerre ; (iii) services de renseignement et sécurité intérieure.

Dans le domaine de l'IA s'est engagée à l'échelle mondiale une course technologique qui défie et met à l'épreuve la stabilité stratégique et la sécurité internationale. Cette évolution englobe également le développement de systèmes d'armes de plus en plus autonomes, qui bouleverseront profondément la conduite de la guerre. Le déséquilibre entre acteurs des secteurs publics et privés soulève aussi des questions relevant de la politique de sécurité. Dans le domaine de l'IA, les grandes entreprises technologiques possèdent une avance considérable en termes de connaissances, de capitalisation de données et d'applications. Enfin, des questions en matière de politique de sécurité se posent dans la surveillance et la lutte contre la criminalité. Là encore, les technologies IA ouvrent de nouvelles perspectives.

6.7.2 Défis

L'IA a des répercussions négatives et positives sur la cybersécurité et la politique de sécurité. L'IA peut être utilisée de manière ciblée pour mener des cyberattaques encore plus rapides et précises (p. ex. espionnage ou phishing), mais aussi à des fins de désinformation et de propagande ou dans les systèmes d'armes. Elle peut aussi servir à renforcer les dispositifs de défense, comme c'est le cas aujourd'hui (p. ex. identification précoce des cybervulnérabilités). De même, les tactiques et vecteurs d'attaque peuvent être identifiés plus rapidement et plus efficacement, et les éléments de sécurité contrôlés et pilotés autant que possible en temps réel (p. ex. protection contre les malwares, pare-feu).

Politique de sécurité extérieure et gouvernance internationale : il convient d'étudier le rôle et l'influence de l'IA sur la politique de sécurité extérieure et les relations internationales. Les questions qui se posent sont les suivantes : (i) dans quelle mesure les systèmes d'IA influencent-ils la stabilité stratégique internationale ? (ii) Dans quelle mesure les systèmes d'IA entraînent-ils une perte de contrôle de l'escalade internationale ? (iii) Quels défis l'IA impose-t-elle au contrôle des armements ?

Forces armées et évolution de la conduite de la guerre : il convient d'anticiper l'influence de l'IA sur les guerres et les conflits, et d'en tirer des conséquences pour nos propres capacités de défense. Les questions qui se posent sont les suivantes : (i) comment l'IA influence-t-elle l'innovation militaire et le développement des capacités ? (ii) Quel est l'impact de l'IA sur les processus de décision militaires et quelles en sont les conséquences (vitesse, gain d'efficience dans l'aide au commandement et l'exploration, implication des parties prenantes, impact sur le droit international humanitaire, appréhension de l'incertitude, solutions stables / instables) ? (iii) Quelles sont les répercussions de l'IA sur les conflits asymétriques ?

Services de renseignement et sécurité intérieure : il convient d'analyser l'influence de l'IA sur les instruments de sécurité intérieure de l'État et d'en évaluer les opportunités et les risques. Les questions qui se posent sont les suivantes : (i) quelles sont les répercussions de l'IA sur l'activité des services de renseignement étatiques ? (ii) Dans quelle mesure l'IA favorise-t-elle l'action de la propagande et de la désinformation ? (iii) Quelle est l'influence des systèmes d'IA sur les activités criminelles dans le cyberespace ? Quelles nouvelles opportunités ou possibilités l'IA offre-t-elle pour les activités de renseignement ?

6.7.3 Activités existantes

Au niveau de la Confédération, il existe actuellement déjà des instruments couvrant et abordant les aspects liés à l'IA et à ses répercussions dans le domaine de la politique de sécurité :

[79] Pour un exposé détaillé, voir le rapport du groupe de projet *Künstliche Intelligenz in der Cybersicherheit und Sicherheitspolitik*, août 2019, disponible à l'adresse www.sbfi.admin.ch/ai-f

Stratégie nationale de protection contre les cyberrisques (SNPC) : la SNPC définit les activités et projets en cours et planifiés contribuant au renforcement de la protection contre les cyberrisques et ayant un lien avec l'IA. Les instruments MELANI (Centrale d'enregistrement et d'analyse pour la sûreté de l'information) et GovCERT (Governmental Computer Emergency Response Team) mettent notamment à la disposition de l'UPIC des capacités d'analyse des nouveaux risques du cyberespace (y compris l'IA).

Stratégie nationale de protection des infrastructures critiques (stratégie PIC) : la stratégie PIC se compose de 17 mesures visant à améliorer la protection des infrastructures critiques et ainsi à garantir la disponibilité de biens et services importants (p. ex. services d'information et de communication). Dans le cadre de cette stratégie, il est également possible de recenser de nouveaux risques liés à l'IA et concernant la disponibilité de services critiques (p. ex. les *Cyber Supply Chain Risks*), tout comme de relever les opportunités offertes pour en améliorer la protection (p. ex. processus de surveillance et de décision basés sur l'IA).

Plan d'action Cyberdéfense (PACD) : avec ce plan d'action en partie confidentiel, le DDPS renforce ses cybercapacités de manière systématique. Outre la propre protection, il s'agit surtout de mettre en œuvre les aspects cyber de la loi sur le renseignement (LRens) et de la loi sur l'armée (LAAM) et d'être en mesure d'assister les opérateurs d'infrastructures critiques subissant des cyberattaques.

Cyberdefence Campus (CYD-Campus) : depuis début 2019, armasuisse S+T gère le CYD-Campus défini dans le cadre du PACD. Le CYD-Campus est une plateforme d'anticipation, d'identification précoce et de veille des nouvelles technologies, dont l'IA. Il s'appuie sur une collaboration étroite avec les hautes écoles (p. ex. les EPF) et l'économie. Le Centre Suisse des Drones et de la Robotique (CSDR) explore, à travers des projets bilatéraux ou multilatéraux, les risques et les opportunités posés par l'action combinée de la robotique et de l'IA pour la sécurité de la Suisse.

Collaboration internationale : le DFAE a renforcé ses activités de politique de sécurité et de sécurité extérieure dans le domaine de la technologie et du cyberespace (p. ex. création de bureaux pour la politique étrangère et de sécurité relative au cyberspace). Sur le plan technique, l'instrument govCERT de l'UPIC entretient des collaborations internationales en vue d'échanger des informations sur la gestion des incidents. Le Service de renseignement de la Confédération entretient des relations étroites avec des services partenaires dans des pays qui sont également concernés par les cyberattaques, soit parce qu'ils sont pris pour cible, soit parce qu'ils hébergent les infrastructures permettant les attaques. L'influence de l'IA est un objet de discussions fréquent lors de ces collaborations aussi brèves que ciblées. L'armée mène des coopérations bilatérales dans le domaine de la cyberdéfense, notamment avec les pays voisins, les échanges portant également sur l'IA. Les coopérations multilatérales sont, elles aussi, renforcées, par exemple avec la participation au Cooperative Cyber Defence Center of Excellence (CCDCOE), à Tallinn, qui traite également de l'IA.

6.7.4 Évaluation et actions requises

<table>
<tr><td colspan="2">Champ d'action 1 : Répercussions sur la politique de sécurité extérieure
L'utilisation de l'IA soulève des questions concernant la stabilité stratégique internationale. L'IA est utilisée comme un instrument d'influence politique et de projection de puissance. Elle influence le déséquilibre entre acteurs des secteurs publics et privés ainsi que la course aux armements.</td></tr>
<tr><td>Examen des implications liés à l'utilisation des systèmes basés sur l'IA pour la politique étrangère</td><td>Le DFAE examine les implications de l'utilisation de systèmes basés sur l'IA pour la politique étrangère. Il en découle des questions relatives à la politique étrangère dans les domaines de la réglementation, du rôle de l'État, des normes et des droits de l'homme. Le contrôle des armements, rendu plus difficile, est également concerné.

Responsabilité : DFAE ; armasuisse (DDPS)
Statut : mise en œuvre dans le cadre des compétences existantes</td></tr>
<tr><td colspan="2">Actions supplémentaires requises : non</td></tr>
</table>

Champ d'action 2 : Formes de menace et doctrine L'utilisation de l'IA soulève des questions quant aux nouvelles formes de menace et à l'accroissement du potentiel de menace des systèmes et solutions dans le domaine militaire et de la sécurité, en particulier en ce qui concerne les infrastructures critiques. L'influence de l'IA et l'évolution résultante de la menace doivent être surveillées et évaluées dans les domaines suivants :	
1) Examen de la cybersécurité dans le cadre des nouvelles formes de menace au vu de l'utilisation de l'IA	Cybersécurité : comment l'IA est-elle utilisée dans le cadre des attaques complexes et comment le niveau de menace s'en trouve-t-il augmenté ? Quels risques découlent de l'emploi de l'IA en lien avec l'interconnexion croissante (Internet des objets) ? **Responsabilité :** SRC ; armasuisse (DDPS) ; Centre de compétences pour la cybersécurité (DFF) ; DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Examen de la propagande et des opérations d'information au vu de l'utilisation de l'IA de l'IA	Opérations d'information et de prise d'influence : comment les campagnes de propagande et de désinformation (p. ex. lors des élections) sont-elles menées ? **Responsabilité :** ChF ; Centre de compétences pour la cybersécurité (DFF) ; SRC ; armasuisse (DDPS) ; DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
3) Examen de la conduite de la guerre et des capacités infraguerrières au vu de l'utilisation de l'IA	Conduite de la guerre : comment les systèmes d'aide à la décision avec IA influenceront-ils la prise de décision militaire ? Comment les acteurs actuels, y compris les spécialistes du droit international humanitaire, seront-ils à l'avenir intégrés dans les processus de décisions ? En quoi les armes de plus en plus autonomes (SALA[80]), à l'interface entre la robotique et l'IA, changeront-elles la conduite de la guerre ? Violences infraguerrières : quelle influence a l'IA sur l'élucidation, le tableau de la situation, la commande et le contrôle (C2I) et l'impact dans des contextes hybrides avec des acteurs non étatiques ? **Responsabilité :** armée ; SG DDPS ; SRC ; armasuisse (DDPS) ; DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 3 : Connaissances et capacités L'utilisation de l'IA influence aussi bien certaines formes de menace récemment apparues que les moyens optimisés de protection et de défense. Les stratégies et plans d'action en place (SNPC, SKI, PACD) doivent tenir compte de l'IA en tant que facteur critique de développement.	
1) Renforcement de l'intégration et de l'utilisation des solutions IA dans les forces armées et le renseignement	Le DDPS procède à l'intégration et à l'utilisation de solutions IA à tous les niveaux des processus militaires (planification, commandement et logistique) et dans les systèmes d'armes (y compris les SALA). Il œuvre en faveur du développement de capacités de protection et de l'agilité par le recours aux technologies IA au sein des instruments de sécurité actuels de la Suisse. **Responsabilité** armée ; SRC ; armasuisse (DDPS) **Statut :** mise en œuvre dans le cadre des compétences existantes

[80] Les systèmes d'armes létales autonomes sont des systèmes développés pour sélectionner et viser des cibles militaires (personnes et installations) sans intervention humaine.

2) Examen des possibilités de mise à niveau dans le cadre du processus d'acquisition de systèmes critiques	Le processus d'acquisition de systèmes critiques doit tenir compte des possibilités de mise à niveau intégrant les fonctionnalités potentielles d'IA à venir. **Responsabilité :** armasuisse ; armée (DDPS) **Statut :** mise en œuvre dans le cadre des compétences existantes
3) Meilleure prise en compte des composants d'IA chez les fournisseurs et sous-contractants de systèmes critiques	La liste de tous les fournisseurs et sous-contractants (*Cyber Security Supply Chain*) de systèmes critiques (de l'armée et de l'administration militaire) doit être établie en tenant compte également des composants IA. **Responsabilité :** SRC ; armasuisse ; armée (DDPS) **Statut :** mise en œuvre dans le cadre des compétences existantes
4) Vérification régulière du niveau technologique des opérateurs d'infrastructures critiques	Le niveau technologique des principaux partenaires (opérateurs d'infrastructures critiques) est contrôlé régulièrement afin de garantir l'interopérabilité dans le cadre des applications IA. **Responsabilité :** armasuisse ; OFPP (DDPS) **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 4 : Anticipation par la collaboration, la recherche et les bancs de test L'étroite collaboration avec les hautes écoles et l'industrie dans la recherche technique et en sciences sociales est un facteur clé pour intégrer avec succès le potentiel de l'IA dans la cybersécurité et la cyberdéfense.	
1) Renforcement de la collaboration avec les grands instituts de formation et de recherche	Dans le cadre de leurs compétences, le DDPS, le DEFR et le DFAE renforcent leur collaboration avec les grands instituts de formation et de recherche et accroissent l'agilité de leur anticipation au moyen de projets de recherche (promotion d'une recherche translationnelle). **Responsabilité :** DDPS, DEFR, DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Prise en compte des développements dans le domaine de l'IA dans l'image de la situation cybernétique	L'image intégrale de de la situation dans le cyberespace tient compte des développements dans le domaine de l'IA. **Responsabilité :** SRC, SG-DDPS, armasuisse (DDPS), Centre de compétences pour la cybersécurité (DFF) **Statut :** mise en œuvre dans le cadre des compétences existantes
3) Renforcement de la participation aux instances et initiatives de recherche internationales dans le domaine de l'IA	La participation ciblée à des instances et initiatives de recherche internationales permet à la Suisse de se positionner comme un partenaire de poids dans le domaine de l'IA. **Responsabilité :** DDPS, DEFR, DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
4) Examen de la nécessité, du rôle et du potentiel d'un banc de test IA pour la Suisse	Le DFF et le DDPS, conjointement avec le DFAE et le DEFR, mènent une étude sur la nécessité pour la Suisse de disposer d'un banc de test IA (cadre actuel, défis à venir, évaluation du besoin d'agir et recommandations, y compris avantages et inconvénients d'un banc de test national). **Responsabilité :** Centre de compétences pour la cybersécurité (DFF), SG-DDPS, armasuisse (DDPS) en collaboration avec le DFAE / DEFR **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.8 Intelligence artificielle, médias et sphère publique[81]

6.8.1 Vue d'ensemble

La sphère publique est la sphère sociale dans laquelle les citoyens échangent sur leurs préoccupations communes[82]. Dans la sphère publique sont diffusées et débattues des informations, valeurs, idées et conceptions concernant la société. Une société démocratique a donc besoin d'une sphère publique qui fonctionne correctement en tant « qu'espace politique » de la société, où des débats sur ses valeurs et ses objectifs, ainsi que les possibilités et les moyens de les atteindre, peuvent se dérouler librement.

Longtemps, le façonnement de la sphère publique et, avec lui, un préalable essentiel à la formation de l'opinion sur les questions politiques, se sont appuyés sur une structure plus ou moins stable de **médias de masse** (télévision, radio et presse écrite). Cette formation de la sphère publique par les médias de masse doit non seulement respecter l'ordre juridique suisse général, mais suit aussi des règles connues, transparentes et négociables au sein de la société : dans l'ensemble du secteur suisse des médias, à travers les « règles déontologiques » définies par la branche elle-même et appliquées par le Conseil suisse de la presse[83] ; dans le secteur de la radio et de la télévision, également au moyen de la loi fédérale du 24 mars 2006 sur la radio et la télévision (LRTV ; RS 784.40).

Au cours des dix dernières années, différentes offres en ligne ont modifié structurellement la formation traditionnelle de la sphère publique : les moteurs de recherche (p. ex. Google), les plateformes de réseaux sociaux (p. ex. Facebook), les plateformes multimédias (p. ex. YouTube) et les services de microblogging (p. ex. Twitter) sont des acteurs de plus en plus importants de la communication publique. Leur point commun est qu'ils ne produisent pas de contenus ou très peu, mais jouent le rôle de médiateurs entre les créateurs de contenus et les consommateurs. C'est pourquoi ces prestataires de services sont dénommés **« intermédiaires »**[84] par les scientifiques. Ils sélectionnent, hiérarchisent, filtrent, agrègent et diffusent l'information. Ils décident par exemple de ce que voient ou ne voient pas les utilisateurs. Comme les médias de masse traditionnels, les intermédiaires exercent donc eux aussi une influence sur la perception du monde, la formation de l'opinion et le comportement humain. Cependant, leurs logiques de sélection reposent sur des règles différentes, qui ne relèvent pas du journalisme et ne sont pas transparentes. Il en découle des opportunités et des risques s'agissant des droits fondamentaux en matière de communication.

Les médias de masse et les intermédiaires font appel à l'intelligence artificielle (IA) notamment pour la sélection et la diffusion des informations. Aujourd'hui, la structuration de la sphère publique, le façonnement de la réalité sociale et la formation de l'opinion des citoyens reposent donc pour une large part sur des services basés sur l'IA.

6.8.2 Défis

Façonnement de la sphère publique par les médias de masse à l'aide de l'intelligence artificielle : le journalisme algorithmique et les applications basées sur l'IA interviennent à toutes les étapes du processus journalistique (agrégation, production et diffusion). Ils peuvent en outre être utilisés pour les contenus de toute nature[85]. En Suisse également, les médias de masse classiques ainsi que l'agence de presse Keystone-SDA utilisent dans leur activité journalistique des logiciels faisant appel aux algorithmes et à l'intelligence artificielle[86]. Dans le domaine de la production de l'information, par exemple, l'agence de presse Keystone-SDA, avec « Lena », et Tamedia, avec « Tobi », se servent de logiciels qui, à l'aide de l'application IA « Natural Language Generation » (NLG), sont capables de rédiger des articles en français et en allemand. On sait aussi que les agences de presse internationales Agence France-Presse (FR), Austria Presse Agentur (AT), PA

[81] Pour un exposé détaillé, voir le rapport du groupe de projet *Intelligence artificielle, médias et sphère publique*, août 2019, disponible à l'adresse www.sbfi.admin.ch/ai-f
[82] Cf. Dreyer et Schulz, 2019, p. 6.
[83] Cf. Conseil suisse de la presse : Code déontologique.
[84] Également appelés « intermédiaires d'informations » ; pour la terminologie, cf. p. ex. Dreyer et Schulz, 2019.
[85] Données, textes, images, audio, vidéo ; cf. pour combinaison des formes d'application Goldhammer et al. 2019.
[86] P. ex. Tamedia, NZZ, Ringier Axel Springer, La Liberté ; cf. Goldhammer et al. 2019, pp. 21–26.

Press Association (GB), Thompson Reuters (GB), Associated Press (US) et Bloomberg (US) ont recours aux algorithmes.

Façonnement de la sphère publique par les intermédiaires à l'aide de l'intelligence artificielle : les intermédiaires (comme les médias de masse classiques) proposent des connaissances disponibles d'une manière qui « peut être variée et objective en théorie, mais aussi partiale et erronée »[87]. Plus leur part sur le marché des utilisateurs est grande, plus les potentiels (négatifs ou positifs selon le cas) des intermédiaires sont importants.

Pour ce qui concerne les risques, on constate que la diversité de l'offre d'informations mise à la disposition d'un utilisateur peut être limitée : soit parce qu'une présélection du contenu est effectuée par « apprentissage automatique », auquel cas l'utilisateur se voit toujours proposer en premier lieu des contenus similaires (*filter bubbles*) ; soit parce que ce type de méthodes permet à l'utilisateur d'exclure plus facilement les opinions indésirables (*echo chambers*). Dans le domaine de la reconnaissance automatique de contenus, par exemple dans la reconnaissance d'images, des contenus licites peuvent être supprimés ou filtrés à la suite d'évaluations incorrectes (« censure »).

Mais les intermédiaires peuvent également intervenir directement dans la formation de l'opinion politique, par exemple en affichant des publicités politiques ciblées anonymes (*dark ads*) ou en expérimentant les incitations au vote[88]. Par ailleurs, les systèmes algorithmiques peuvent être exploités par des acteurs tiers « externes » : Par exemple, les *social bots* peuvent être utilisés pour tirer parti des effets multiplicateurs des plateformes en vue de renforcer des intérêts particuliers ou d'affaiblir la visibilité d'opinions divergentes[89]. Le Conseil fédéral a alerté sur cette question dans son rapport *Cadre juridique pour les médias sociaux*[90]. Dans ce contexte, citons également les *shit storms* et les *hate speech*[91].

Il serait certainement faux de supposer que les fausses informations diffusées via les médias sociaux puissent modifier immédiatement l'opinion, voire le comportement (en matière d'élection et de votation), de leurs destinataires. Et dans le cas de la formation de l'opinion par les intermédiaires comme de celle passant par les médias, la méfiance vis-à-vis des élites, le mécontentement envers la politique, les inégalités économiques ou l'exclusion culturelle exercent sans doute une influence plus forte sur les choix électoraux individuels[92]. De plus, les études empiriques montrent que, globalement, l'utilisation des médias sociaux favorise plutôt la diversité des discours à l'heure actuelle. Il ne semble pas non plus que « l'agenda des grands sujets de société » soit fragmenté par les intermédiaires, et il n'y a actuellement pas d'élément valable attestant l'existence de bulles de filtres[93].

Cependant, il est évident que les intermédiaires ont (en théorie) le potentiel d'instrumentaliser les applications IA à des fins commerciales ou politiques ou d'êtres eux-mêmes instrumentalisés à ces mêmes fins. De ce fait, la formation de l'opinion et de la volonté publiques peut être influencée, y compris dans le domaine politique[94]. En règle générale, les intermédiaires ne poursuivent pas d'objectifs d'intérêt général (p. ex. la diversité des opinions), mais servent des intérêts économiques particuliers. En outre, leur puissance de communication renferme également des risques pour l'ouverture des processus de communication au sein de la société[95]. Le Conseil fédéral s'est déjà exprimé dans le passé sur la responsabilité des intermédiaires. La question de la mise en œuvre du droit dans le contexte des intermédiaires fait notamment l'objet d'interventions transmises[96]. Cependant, la question essentielle de la responsabilité juridique et sociale des intermédiaires reste en

[87] Dreyer et Schulz 2019, p. 7, cf. également ; Lobigs et Neuberger 2018.
[88] Cf. Fichter 2018.
[89] Cf. Gillespie 2017.
[90] Cf. Conseil fédéral 2017.
[91] Cf. Jarren 2018b, p. 36.
[92] Cf. Livingstone 2019 ; Commission fédérale des médias COFEM 2019.
[93] Cf. Dreyer et Schulz 2019, 11 ; 16-19.
[94] Le présent rapport se concentre sur les activités et les potentiels des intermédiaires qui ont un lien direct avec le façonnement de la sphère publique. Les autres thèmes en rapport avec les activités des intermédiaires, par exemple l'utilisation des données, ne sont pas abordés ici.
[95] Cf. Saurwein et al. 2017.
[96] Cf. récemment le rapport du Conseil fédéral *Responsabilité civile des fournisseurs Internet* du 11 décembre 2015, disponible à l'adresse : <https://www.ejpd.admin.ch/ejpd/fr/home/aktuell/news/2015/2015-12-110.html>, ainsi que les motions 18.3379 CAJ-E « Accès des autorités de poursuite pénale aux données conservées à l'étranger » et 18.3306 Glättli « Renforcer l'application du droit sur Internet en obligeant les grandes plates-formes commerciales à avoir un domicile de notification ».

suspens. Eu égard aux changements rapides que connaît le processus de façonnement de la sphère publique et du recours croissant aux intermédiaires par la population, cette question est de plus en plus d'actualité et gagne en importance.

6.8.3 Activités existantes

Façonnement de la sphère publique par les médias de masse à l'aide de l'intelligence artificielle : en l'état actuel des connaissances, il n'y a en Suisse aucune activité réglementaire spécifique ni demande en ce sens de la part des acteurs de la société dans le domaine des médias de masse et de l'intelligence artificielle.

Façonnement de la sphère publique par les intermédiaires à l'aide de l'intelligence artificielle : plusieurs pays adoptent des réglementations sur les intermédiaires. En Allemagne, les « plateformes de médias » et les « intermédiaires de médias » doivent être prochainement inclus dans le traité médias d'État et les robots sociaux soumis à une obligation de signalement. La France a adopté une « loi contre la manipulation de l'information ». Au Royaume-Uni, une loi relative aux réseaux sociaux visant à garantir « un Internet sûr » est en projet.

Dans son rapport consacré au cadre juridique pour les médias sociaux publié en 2017, le Conseil fédéral estime « qu'il n'est actuellement pas nécessaire de prendre des mesures de réglementation supplémentaires en ce qui concerne les médias sociaux »[97]. Certes, il reconnaît le risque d'influence des fausses informations et surtout des robots sociaux sur la formation démocratique de l'opinion. Cependant, faute de recul, il ne souhaite pas pour l'heure répondre à la question de la nécessité d'une réglementation étatique. Il préfère miser sur une autorégulation de la branche et observer attentivement l'évolution de la situation au niveau national et international[98].

6.8.4 Évaluation et actions requises

Champ d'action 1 : Gouvernance suisse dans le domaine des intermédiaires Étant donné l'influence importante des intermédiaires, il convient d'examiner la question de manière approfondie et d'élaborer une approche de la gouvernance suisse.	
Rédaction d'un rapport de gouvernance dans le domaine des intermédiaires	Un rapport de gouvernance examinant et, le cas échéant, proposant des mesures devra être présenté au Conseil fédéral au printemps 2021. **Responsabilité :** OFCOM/ChF **Statut :** examen à l'attention du Conseil fédéral
Actions supplémentaires requises : oui	

Champ d'action 2 : Observation de l'évolution de la situation dans le domaine des médias L'utilisation de l'IA dans le domaine des médias soulève des questions quant à l'explicabilité, à la transparence et à la responsabilité.	
Suivi de l'évolution de l'utilisation de l'IA dans le domaine des médias	Les approches réglementaires des autres pays (p. ex. projet de « traité médias d'État » des *Länder* allemands) doivent également être suivies, de même que les débats scientifiques sur l'explicabilité / la transparence et la responsabilité dans le cadre de l'utilisation de l'IA dans le domaine des médias. Cela peut se faire dans le cadre des activités existantes de l'administration. **Responsabilité :** OFCOM / DFAE **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

[97] Conseil fédéral 2017, p. 52.
[98] ibid, p. 52.

6.9 Mobilité automatisée et intelligence artificielle[99]

6.9.1 Vue d'ensemble

L'utilisation de l'intelligence artificielle (IA) dans la mobilité automatisée permet de mieux exploiter le potentiel important de l'automatisation[100] en vue d'améliorer la sécurité routière et d'accroître l'efficacité du système de transport, par exemple en termes de capacité, de taux d'occupation, de durabilité et de financement. Cela vaut aussi bien pour les moyens de transport eux-mêmes que pour les infrastructures de transport de manière générale, ainsi que pour les systèmes centralisés qui en dépendent. L'utilisation de l'IA optimise un grand nombre de moyens de transport connectés (automatisés) sur le plan du temps, du confort, des coûts, de l'environnement ou des expériences vécues. Elle permet de répondre aux préférences individuelles comme aux objectifs de la société. L'IA est déjà utilisée dans la mobilité, par exemple dans les systèmes d'assistance à la conduite des véhicules pour la reconnaissance de l'environnement ou dans les infrastructures de transport pour la mesure de la fluidité du trafic. Cependant, un véhicule automatisé uniquement au moyen d'algorithmes déterministes ne peut pas réagir à des événements imprévus dans des situations complexes, par exemple un carrefour en centre-ville où se croisent tous les usagers possibles des moyens de transport individuels et des transports publics. Des algorithmes d'apprentissage automatique sont nécessaires.

6.9.2 Défis

Sécurité du trafic et de l'exploitation dans un système global de transport basé sur l'IA : dans un système global de transport connecté et automatisé, le niveau de sécurité est influencé non seulement par les moyens de transport proprement dits, mais aussi par d'autres systèmes techniques et processus d'exploitation. En conséquence, des mesures de sécurité complètes doivent également être prises pour l'ensemble du système, ce qui impose une évaluation globale de la sécurité. Les futurs rôles et tâches de l'État et des acteurs privés – et donc leur responsabilité respective – restent à être déterminés, notamment s'agissant de l'homologation des moyens de transport automatisés ou de la cybersécurité.

L'une des conditions essentielles au bon fonctionnement des systèmes de transport automatisés est que l'IA puisse améliorer ses performances de manière systématique au moyen de données proches de la réalité. La définition de règles relatives à la création, à la collecte et à la mise à disposition de telles données pourra à l'avenir permettre et faciliter leur accès ainsi que renforcer la protection contre les manipulations. La question de savoir si l'on doit conférer aux moyens de transport automatisés pilotés par un système d'IA la même tolérance aux erreurs qu'à ceux commandés par l'homme est loin d'être résolue au sein de la société et est confrontée à des obstacles politiques.

Protection des données et application des dispositions légales dans le transport : l'IA offre de nombreuses possibilités d'analyse des données utilisées dans le domaine de la mobilité, le plus souvent disponibles sans coût supplémentaire, qui consignent et enregistrent presque toutes les activités des utilisateurs. Cela soulève des questions quant aux atteintes à la liberté individuelle et à la protection des données. Les réponses à ces questions ne doivent pas être apportées par le marché, mais faire l'objet d'un vaste débat de société. Par ailleurs, il convient de clarifier dans quelle mesure l'État doit ou peut utiliser ces données aux fins de l'application des dispositions légales.

Gestion efficace du trafic et de la mobilité : on suppose qu'une IA commandera les moyens de transport à l'avenir. Un équilibre social, politique et économique devra être trouvé concernant les règles selon lesquelles l'IA devra optimiser à la fois les souhaits individuels des usages et les besoins collectifs de la société. La mobilité automatisée connectée entraîne une augmentation exponentielle des quantités de données disponibles : en raison, d'une part, des hautes performances des systèmes de contrôle basés sur l'IA et, d'autre part, de l'apparition de nouveaux services de mobilité comme le

[99] Pour un exposé détaillé, voir le rapport du groupe de projet *Automatisierte Mobilität und künstliche Intelligenz*, août 2019, disponible à l'adresse : www.sbfi.admin.ch/ai-f

[100] En Suisse et en Europe, les autorités emploient le terme « automatisé » et non « autonome » pour souligner que les moyens de transport automatisés doivent être connectés. Voir aussi le rapport du Conseil fédéral Conduite automatisée de décembre 2016 ou *Europe on the Move III*, notamment @COM/2018/283.

sharing. Les débats portant sur la protection des données et les droits d'utilisation de ces données devraient donc s'intensifier.

Statut juridique des systèmes automatisés dans le domaine de la mobilité : les systèmes de mobilité automatisée auront également des conséquences négatives. Par exemple, les véhicules d'automatisation élevée ou complète seront impliqués dans des accidents. Le Conseil fédéral a déjà examiné ces questions et estime qu'aucune mesure n'est nécessaire à l'heure actuelle (cf. chapitre 4.2) [101]. Il convient en revanche de continuer de suivre attentivement l'évolution de la situation, en particulier dans les autres pays.

6.9.3 Activités existantes

L'**OFROU**, pour le trafic routier, et l'**OFT**, en trafic ferroviaire, suivent l'évolution de la situation dans le domaine de la mobilité automatisée, publient des rapports sur le sujet et autorisent les essais. En se fondant sur les connaissances disponibles, ils évaluent les besoins en termes de réglementation afin de permettre le développement de la mobilité automatisée, tout en en contrôlant les risques. L'**OFCOM** assure la stabilité et la performance des infrastructures de télécommunications, qui sont indispensables aux systèmes de mobilité fondés sur l'IA. L'OFCOM suit l'évolution de la situation au niveau national et international, notamment dans le domaine de « l'Internet of Things » (IoT). L'OFT met en œuvre ses plans de mesures en faveur des services de mobilité multimodale. L'**OFAC** traite les défis découlant de l'utilisation d'aéronefs sans occupants.

6.9.4 Évaluation et actions requises

D'une manière générale, les défis décrits sont similaires à ceux posés par la numérisation. Toutefois, l'utilisation de l'IA dans la mobilité automatisée les renforce. Ils sont actuellement sujets à controverse parmi toutes les parties prenantes. Les pouvoirs publics doivent mobiliser des ressources supplémentaires pour ne pas rester à la traîne de ces développements. Dans l'intérêt de l'ensemble de la société, les opportunités offertes par les nouvelles technologies doivent être saisies et les risques afférents minimisés.

Les offices compétents sont sensibilisés aux questions fondamentales que l'utilisation de l'IA dans la mobilité automatisée devrait soulever. Différents travaux sont actuellement menés sur le sujet. Aucune action supplémentaire n'est en principe requise concernant l'IA. De nouvelles actions pourraient en revanche devenir nécessaires à la suite des mesures et des développements en cours.

Champ d'action 1 : Utilisation de l'IA dans les véhicules automatisés L'utilisation de l'IA dans les véhicules automatisés est nécessaire pour qu'ils puissent circuler en toute sécurité sur les infrastructures existantes en même temps que les véhicules conventionnels (commandés par l'homme).	
1) Coordination des travaux en cours portant sur les véhicules automatisés	L'OFROU et l'OFT suivent l'évolution de la situation au niveau national et international et coordonnent les travaux sur les véhicules automatisés (voir rapport du Conseil fédéral *Conduite automatisée – Conséquences et effets sur la politique des transports* [décembre 2016]). **Responsabilité :** OFROU / OFT **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Définition de la gestion du trafic aérien pour les systèmes aériens sans pilote (*Unmanned Aircraft Systems, UAS*)	Définition de règles spécifiques de gestion du trafic aérien pour les systèmes aériens sans pilote (*Unmanned Aircraft Systems, UAS*) **Responsabilité :** OFAC **Statut :** mise en œuvre dans le cadre des compétences existantes

[101] Réponses aux interventions 18.3445 Ip. Marchand-Balet « Véhicules autonomes et responsabilité. À quand une adaptation de la législation helvétique ? » et 17.3040 Po. Reynard « Examen de la création d'une personnalité juridique pour les robots ».

3) Éclaircissements concernant un projet pilote de conduite automatique en trafic ferroviaire et routier	Éclaircissements concernant un projet pilote de conduite automatique (intelligence artificielle dans les centrales d'exploitation / en gestion du trafic). **Responsabilité :** OFT / OFROU **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 2 : Échanges de données obligatoires pour l'IA dans la mobilité automatisée Les échanges de données entre tous les usagers des transports et l'infrastructure sont intensifiés afin de garantir le fonctionnement optimal de l'IA dans la mobilité automatisée.	
1) Mise en œuvre des plans de mesures existants	Mise en œuvre des plans de mesures existants, notamment en faveur des services de mobilité multimodale et création d'une plateforme de données de trafic *(*voir rapports *Prestations de mobilité multimodale ; Plans de mesures : données mobilitaires et ouverture de la distribution des fournisseurs de mobilité externes aux TP* [décembre 2018] et *Mise à disposition et échanges de données pour la conduite automatisée dans le trafic routier* [décembre 2018]). **Responsabilité :** OFT/OFROU **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Développement du « réseau de transport suisse » de géolocalisation	Développement d'un « réseau de transport suisse » de géolocalisation. **Responsabilité :** swisstopo **Statut :** mise en œuvre dans le cadre des compétences existantes
3) Clarifications concernant le projet pilote de communication mobile de sécurité à large bande (CMS)	Clarifications à propos de l'éventualité d'un projet pilote de communication mobile de sécurité à large bande (CMS) **Responsabilité :** DDPS (OFPP) **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 3 : Protection des données dans la mobilité automatisée Les nombreuses possibilités d'utilisation de l'IA dans la mobilité au sens large sont préservées dans le respect des dispositions relatives à la protection des données.	
Réseau de coordination avec le PFPDT	La Confédération coordonne un suivi réglementaire étroit. L'OFROU et l'OFT définissent les mesures à prendre, notamment en ce qui concerne la réalisation des objectifs en matière de mobilité et l'application des prescriptions légales. **Responsabilité :** OFROU/OFT **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d’action 4 : **Législation et acceptation sociale de l’IA dans la mobilité automatisée** Création d’un cadre juridique pour l’IA dans la mobilité automatisée et promotion de son acceptation sociale.	
1) Autorisations pour la conduite automatisée	Les autorisations pour la conduite automatisée contribuent à l’acceptation de l’IA dans la mobilité automatisée et permettent l’acquisition de connaissances, notamment en vue de la révision de la loi sur la circulation routière (LCR) et de la loi fédérale sur les chemins de fer (LCdF). **Responsabilité :** OFROU/OFT **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Coordination des procédures d’homologation des véhicules automatisés	Les services du DETEC coordonnent les procédures d’homologation des véhicules automatisés. **Responsabilité :** DETEC **Statut :** mise en œuvre dans le cadre des compétences existantes
3) Clarification des tolérances aux erreurs de l’IA	L’OFROU et l’OFT définissent les tolérances aux erreurs pouvant être acceptées en transports automatisés dans le domaine de l’IA . **Responsabilité :** OFROU/OFT **Statut :** mise en œuvre dans le cadre des compétences existantes
4) Réseau de coordination dans le domaine juridique et international	Les services du DETEC garantissent la coopération juridique et internationale. **Responsabilité :** DETEC **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.10 L'intelligence artificielle dans la santé

6.10.1 Vue d'ensemble

La numérisation modifie également le secteur de la santé à un rythme fulgurant : il n'y a jamais eu autant de moyens de collecter, de compiler et d'analyser les données de santé. La puissance de calcul exponentielle et les nouvelles techniques de recoupement et de traitement de l'information facilitent considérablement l'analyse de ces données. On attribue un grand potentiel aux données de santé qui, si elles sont correctement extraites, peuvent non seulement être utiles pour la recherche médicale, mais aussi contribuer à la fourniture de soins efficaces et optimaux ainsi qu'à l'amélioration de la santé publique. L'intelligence artificielle (IA) et les algorithmes toujours plus complexes offrent des opportunités importantes au secteur de la santé. Des améliorations sont annoncées à tous les niveaux grâce aux différentes méthodes de l'IA : qualité des soins, offre de prestations de santé, intérêt pour le patient, efficacité des coûts.

6.10.2 Défis

Les défis résident tout d'abord dans l'évaluation correcte des opportunités que la numérisation offre au secteur de la santé. Par ailleurs, les risques possibles doivent être identifiés précocement et éliminés ou, au moins, minimisés.

Les *opportunités* offertes par le traitement et l'analyse des données de santé, la mise en pratique des informations recueillies et le développement des concepts et approches de la médecine axée sur les données sont les suivantes :

- Amélioration de la prévention et de la promotion de la santé grâce à une identification plus précoce et plus complète des divers facteurs de risque et à l'élaboration de mesures visant à modifier le comportement en matière de santé.
- Soins plus efficaces et davantage centrés sur le patient par le remplacement des traitements inefficaces, voire préjudiciables par de nouvelles thérapies personnalisées (p. ex. en oncologie).
- Amélioration du suivi épidémiologique et de la prévision des épidémies.
- Accroissement de la sécurité des médicaments grâce à l'amélioration de la surveillance et à une réduction des effets secondaires permise par une utilisation ciblée des médicaments et de nouvelles méthodes de diagnostic.
- Amélioration du rapport coût-efficacité dans le système de santé grâce à l'identification des offres excédentaires et à la mise en place d'une offre de soins adaptée aux besoins ainsi qu'à l'accroissement de l'efficacité.

Les *risques* les plus fréquents de la médecine axée sur les données concernent la protection des données et de la personnalité. Plus les fichiers contenant des données personnelles sont volumineux, plus il est difficile de les anonymiser. Bien que les résultats des recherches menées à l'aide de ces données ne permettent généralement pas d'identifier les personnes concernées, il existe toujours un risque que des données personnelles soient divulguées en raison de failles de sécurité ou à la suite d'un acte délibéré. De surcroît, du fait du développement rapide des techniques d'analyse (p. ex. le séquençage à haut débit du génome), des nouvelles technologies numériques et des possibilités de mise en relation des données, on ne peut pas exclure que l'anonymisation et le chiffrage des données puissent être facilement contournés. Une protection absolue des données est impossible sur le plan technique. La transmission involontaire de données de santé renferme des risques qui touchent principalement à l'identité, à la sphère privée, aux droits de propriété intellectuelle et de la personnalité et à la discrimination.

Certains facteurs peuvent en outre avoir un impact négatif sur la qualité / l'exactitude des données, ce qui présente également des risques. Dans le cas de l'intelligence artificielle ou des systèmes d'apprentissage automatique utilisés dans les analyses de données, les algorithmes et les règles de décision ont souvent été élaborés de manière évolutive et ne sont pas soumis à une analyse ou à une vérification extérieure (boîte noire). Il est par conséquent difficile de s'assurer de l'exactitude d'une information générée par ce biais par l'analyse des modalités de sa formation. Les systèmes auto-apprenants peuvent également « mal » apprendre.

6.10.3 Activités existantes

Avec la médecine axée sur les données, l'OFSP est tiraillé entre deux objectifs. D'un côté, il doit veiller à la protection de la personnalité et du droit à l'autodétermination des patients et des personnes bien portantes. De l'autre, les priorités du Conseil fédéral en matière de politique de santé incluent l'amélioration de la qualité des soins et le maintien de leur accessibilité par des gains d'efficience. Sur toutes ces questions, la numérisation en général, la médecine axée sur les données et l'intelligence artificielle offrent des opportunités importantes, mais renferment aussi des risques. L'OFSP reconnaît qu'une amélioration générale de la standardisation et de la disponibilité des données de santé améliorerait et soutiendrait la recherche s'appuyant sur des données de santé à caractère personnel. L'utilité de ces données pour les soins et la politique de données pourrait être également accrue. Mais, dans le même temps, l'autodétermination en matière d'information des personnes concernées doit être renforcée, de même que la protection et la sécurité des données.

6.10.4 Évaluation et actions requises

À l'avenir, les défis susmentionnés devraient de plus en plus concerner le système et la politique de santé de la Suisse. Dans ce contexte, l'OFSP observera les répercussions de l'IA sur la médecine et les soins et, le cas échéant, présentera au Conseil fédéral des propositions d'adaptation des bases légales du droit fédéral applicables.

Champ d'action 1 : Recherche sur l'être humain	
1) Examen des bases légales dans le domaine de la réutilisation des essais et des données, ainsi que des biobanques	Dans le cadre de l'évaluation de la loi relative à la recherche sur l'être humain (LRH, fin 2019), il sera examiné si les bases légales relatives à la réutilisation des essais et des données sont adaptées au vu des développements actuels et si, pour ce qui concerne les biobanques, la protection des personnes concernées, la liberté scientifique et la santé publique sont garanties. **Responsabilité :** OFSP **Statut :** mise en œuvre dans le cadre des compétences existantes
2) Positionnement cohérent dans le domaine de l'utilisation des données et du *Big Data*	Dans le domaine de l'utilisation des données de santé dans la recherche (en particulier dans le domaine du *Big Data*), l'OFSP se positionnera plus clairement et se concertera avec d'autres organismes fédéraux (p. ex. dans le cadre de la stratégie « Suisse numérique ») pour que les droits des patients et citoyens en matière de sphère privée et de contrôle des données soient préservés autant que possible et qu'ils puissent décider de leur utilisation en toute connaissance de cause. **Responsabilité :** OFSP **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 2 : Loi sur les produits thérapeutiques (LPTh) L'IA ne pourra déployer pleinement son potentiel pour l'amélioration des soins que si elle est utilisée non seulement dans le cadre de projets de recherche, mais aussi dans le processus clinique. La LRH couvre seulement une partie de la problématique. La LPTh deviendra sans doute plus importante à long terme.	
Examen de futures approches dans le développement de médicaments	L'OFSP et Swissmedic examinent les approches tournées vers l'avenir permettant de répondre de manière adéquate à la tendance en faveur de la médecine de précision dans le développement de médicaments. **Responsabilité :** OFSP/Swissmedic **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.11 L'intelligence artificielle dans la finance

6.11.1 Vue d'ensemble

Si le secteur financier emploie depuis longtemps déjà les méthodes de l'intelligence artificielle, par exemple pour identifier les transactions illégales par carte de crédit, on observe actuellement une extension des domaines d'application[102]. L'intérêt de l'IA est qu'elle entraîne une baisse des coûts du fait que les tâches à forte intensité de main-d'œuvre peuvent, dans une très large mesure, être automatisées ou accélérées. Par ailleurs, elle peut permettre de créer des produits qui paraissent plus utiles, plus simples, plus avantageux ou plus personnalisés au client. Des connaissances plus précises sur les clients peuvent se traduire par des décisions précises. L'IA devrait améliorer l'allocation des ressources, car chaque centime non alloué à un projet de moindre intérêt est disponible pour un projet plus prometteur.

6.11.2 Défis

Du fait des dispositions sur la protection du client et des exigences de stabilité, le secteur de la finance est plus réglementé que beaucoup d'autres branches. La probabilité que les applications d'une nouvelle technologie soient en contradiction avec la réglementation en vigueur ou doivent s'y adapter est donc plus élevée. Il en résulte que l'utilisation de l'IA dans la finance doit être suivie de plus près que dans d'autres secteurs économiques. Le phénomène général qui veut que les variantes plus complexes des méthodes IA peuvent difficilement expliquer les résultats dans les cas individuels s'observe également dans les applications du secteur financier. La conséquence peut en être que, premièrement, les erreurs de décision d'un algorithme ne seront pas identifiées et que, deuxièmement, un prestataire de services financiers ne sera pas en mesure de respecter une obligation légale de comportement, de responsabilité, d'informer ou de rendre compte[103] vis-à-vis d'un client.

Si les données utilisées dans la phase d'apprentissage d'une application IA reflètent une discrimination, le système la reprendra et l'appliquera. Une discrimination peut être fondée sur des décisions individuelles, mais aussi sur d'autres artefacts, qui sont parfois aléatoires. Dans certains cas, seul l'établissement en supporte les répercussions négatives, par exemple s'il en résulte des investissements trop élevés ou trop faibles dans certaines catégories de placement (mauvaise allocation). Il serait en revanche significatif sur le plan réglementaire si un prestataire de services financiers conseillait mal un client, ne réalisait pas correctement la vérification de l'adéquation et du caractère approprié imposée par la loi ou désavantageait un assuré sans motif d'ordre juridique ou actuariel[104].

Le fait que les algorithmes d'IA puissent aboutir à des prédictions et à des résultats différents[105] dès qu'une modification minime imperceptible est apportée aux données d'entrée peut être utilisé de manière abusive. Cela offre la possibilité à des tiers de modifier le comportement des systèmes d'IA des établissements financiers au moyen d'une manipulation des données à peine visible pour l'homme. On ne peut pas exclure que l'IA soit un jour utilisée, par des modifications non identifiées des données, à des fins frauduleuses dans le but, par exemple, d'octroyer des crédits abusifs, de débloquer des fonds de sponsoring, d'exécuter des virements, de comptabiliser incorrectement des opérations, de sous-estimer des risques d'assurance et d'identifier de manière erronée des clients

[102] P. ex. identification du client ; classification des e-mails ; *chatbot* de service à la clientèle ; proposition personnalisée de produits de placement, bancaires ou d'assurance ; affectation du client à une catégorie de services ; établissement individualisé de la solvabilité du client ; prime d'assurance ; prévision au jour près des mouvements de compte ; recherche de passages pertinents dans des documents ; prévisions de cours ; identification des transactions illégales sur la base d'anomalies.

[103] P. ex. l'obligation visée à l'art. 15, al. 2 de documenter et de communiquer au client les motifs sous-jacents de chaque recommandation d'acquisition ou d'aliénation d'un instrument financier. Voir aussi le droit à l'explication fixé au considérant 71 relatif à l'article 22 du règlement général sur la protection des données de l'UE.

[104] Loi sur les services financiers (LSFin), art. 6 à 16. Ordonnance sur la surveillance (OS), art. 117.

[105] Su, Vargas et Kouichi. *One pixel attack for fooling deep neural networks*, 2017, arXiv:1710.08864 [cs.LG]. https://arxiv.org/abs/1710.08864

dans les transactions électroniques. Il s'agirait en l'occurrence d'un risque opérationnel relevant du droit de la surveillance.

Le conflit entre la protection des données clients à caractère personnel et les données requises par les systèmes d'IA pourrait être moins important dans les banques que dans le secteur des assurances. En effet, si les produits bancaires sont de plus en plus personnalisés, ils s'appuient toutefois moins sur les données personnelles que les produits d'assurance. En présence d'un dispositif souple de protection des données ou d'une communication volontaire de données, il est possible qu'à l'avenir, les primes des assurés soient par exemple fonction de leurs achats de produits alimentaires et de livres, de leur utilisation d'Internet et des appareils électroniques, de leurs habitudes de voyage, de leur trajet quotidien domicile-travail et de nombreux autres facteurs[106].

Associée aux importants volumes de données, l'IA permet aux assureurs de distinguer avec plus de précision les risques élevés des risques faibles. Il en résulte des primes d'assurance dites conformes au risque, calculées en fonction des caractéristiques spécifiques du risque à assurer. Pour ce faire, les compagnies d'assurances utilisent certes des méthodes statistiques éprouvées depuis longtemps, mais seule l'IA permet d'étayer le tarif des primes sur un grand nombre de caractéristiques de données et leurs relations non linéaires. Les *assurances sociales*, par exemple l'assurance obligatoire des soins selon la LAMal, reposent sur le principe de solidarité entre les assurés concernant le montant des primes[107]. Une prime qui serait uniquement établie sur la base du risque ne serait donc pas socialement acceptable. Dans le domaine des *assurances privées*, hormis les risques dans la prévoyance professionnelle et l'assurance-maladie complémentaire, l'État n'intervient pas dans l'établissement des primes d'assurance. Du point de vue du droit de la surveillance, rien ne s'oppose à un prix individuel pour chaque risque pour autant qu'il puisse être déterminé correctement. Il est à noter que l'individualisation des primes ne porte *pas* préjudice au concept d'assurance de compensation collective des risques. La diversification des risques à l'intérieur de la communauté de risques reste entière, car elle ne résulte pas du nivellement des primes, mais de l'indépendance et de la contingence de la survenance des sinistres pour les différents risques assurés.

Les assureurs n'entrent en contact avec l'IA pas seulement en tant qu'utilisateurs, mais aussi à travers l'assurance d'entreprises qui utilisent l'IA ou commercialisent des systèmes d'IA, ainsi que l'assurance de produits dotés d'une IA. Certaines des questions juridiques (relevant du droit de la responsabilité civile) évoquées en introduction se posent ici également.

6.11.3 Activités existantes

Le DFF suit les développements de l'IA dans le secteur financier, mais n'a jusqu'à présent pas jugé nécessaire d'adopter des mesures réglementaires spécifiques à l'IA. Conformément aux lois sur les marchés financiers, l'autorité de surveillance des marchés financiers surveille les assujettis, par exemple les compagnies d'assurances et les banques.

6.11.4 Évaluation et actions requises

Le législateur et l'autorité de surveillance devraient s'attaquer, ne serait-ce que ponctuellement, aux défis mentionnés, par exemple le faible niveau d'explicabilité, certaines formes d'une possible discrimination, le risque opérationnel représenté par le potentiel d'abus au moyen de données d'entrée modifiées, la protection des données, en particulier dans le secteur des assurances. Dans ce contexte, le DFF suivra les domaines d'application de l'IA chez les prestataires de services financiers et traitera les questions qui se feront jour dans le cadre des révisions ordinaires de la réglementation.

[106] Produits alimentaires : un assuré qui règlerait ses légumes en espèces et ses friandises avec un moyen de paiement électronique pourrait voir sa prime d'assurance augmenter, car le système IA de l'assureur n'enregistrerait que l'achat de friandises. / Livres : les habitudes de lecture peuvent par exemple révéler des informations sur le statut social, la solvabilité et la personnalité. / Téléphones portables : (i) les données de l'accéléromètre et du gyroscope du smartphone peuvent révéler si l'utilisateur emprunte les escaliers ou l'ascenseur. (ii) Les données de mouvement peuvent pointer un manque de sommeil. (iii) En raison de l'altitude, les passagers aériens sont mal protégés contre les rayons cosmiques. Les assureurs pourraient établir une relation entre cette information et la santé de leurs clients.

[107] Par exemple parce que la société souhaite que tous ses membres bénéficient d'une couverture d'assurance ou que certaines activités, régions ou catégories de la population présentant un risque accru doivent être protégées par l'État.

Une modification de la réglementation des marchés financiers ciblant spécifiquement l'IA n'est pas prévue à l'heure actuelle.

Champ d'action 1 : Obligations de comportement Il n'est pas exclu que les futures applications de l'IA utilisées dans le secteur financier soient en contradiction avec les obligations de comportement imposées aux acteurs des marchés financiers par la réglementation actuelle.	
Suivi des développements concernant les obligations de comportement	Le DFF suit les développements de l'utilisation de l'IA dans le secteur financier ainsi que l'éventuelle nécessité d'adaptation des obligations de comportement. Pour l'heure, aucune mesure n'est prévue. **Responsabilité :** DFF **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 2 : Risques opérationnels Les applications de l'IA utilisées dans les établissements financiers peuvent entraîner des risques opérationnels relevant du droit de la surveillance.	
Suivi de l'évolution des risques opérationnels dans les établissements financiers	Le DFF suit les développements des applications IA. La réglementation décrit les risques opérationnels dans des termes suffisamment généraux. Une modification de la réglementation ne paraît donc pas nécessaire. **Responsabilité :** DFF **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 3 : Fixation des primes d'assurance des compagnies d'assurance privées Une utilisation inconsidérée des méthodes de l'IA et des artefacts dans les données pourrait entraîner des discriminations injustifiées sur le plan actuariel dans la structure des primes.	
Suivi des développements concernant les primes d'assurance des compagnies d'assurance privées	Dans le secteur des assurances privées, les primes relatives aux risques dans la prévoyance professionnelle et l'assurance-maladie complémentaire à l'assurance-maladie sociale sont contrôlées par l'État. L'autorité de surveillance a également pour mission de protéger les assurés des abus. Une adaptation de la réglementation ne paraît pas nécessaire. **Responsabilité :** DFF **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.12 L'intelligence artificielle dans l'agriculture

6.12.1 Vue d'ensemble

L'IA est également de plus en plus présente dans l'agriculture. L'amélioration de la précision des technologies informatiques cognitives telles que la reconnaissance d'images transforme l'agriculture, un secteur qui reposait traditionnellement sur l'expérience des agriculteurs afin de connaître par exemple le meilleur moment pour cueillir les fruits. Ainsi, des robots de récolte équipés de la technologie IA peuvent prendre des décisions en temps réel concernant les tâches à accomplir de manière autonome – en s'appuyant sur les données issues de caméras et de capteurs. Les robots de ce type sont de plus en plus capables d'effectuer des tâches agricoles (même complexes) qui nécessitaient auparavant une intervention humaine.

Divers instituts de recherche et des start-up technologiques étudient et développent également en Suisse des solutions innovantes visant à accroître l'utilisation de l'IA dans l'agriculture. Dans ce cadre, la technologie IA est notamment mise en œuvre dans les domaines d'application suivants :

Robots agricoles : ces robots exécutent des tâches agricoles essentielles pour soulager les agriculteurs, par exemple la récolte des plantes utiles ou la lutte contre les plantes nuisibles. Ces robots (auto-apprenants) deviennent sans cesse plus rapides et plus productifs que la main-d'œuvre humaine. Par exemple, la start-up suisse ecoRobotix a mis au point une première machine entièrement autonome basée sur la reconnaissance d'images permettant un désherbage plus économe et plus respectueux de l'environnement.

Surveillance des plantes utiles et des sols : on fait appel à des algorithmes des domaines *computer vision* et *deep learning* pour collecter et traiter des données sur l'état des plantes utiles et des sols. La disponibilité croissante des données satellitaires améliore sans cesse la surveillance. Pour l'agriculteur, l'intérêt réside par exemple dans une utilisation plus ciblée de l'eau, des semences, des engrais et des produits phytosanitaires ainsi que dans la fixation de la date de récolte optimale. Un projet de recherche de l'ETH Zurich portant sur l'état des surfaces agricoles suisses associe la télédétection à l'apprentissage automatique en interprétation d'images.

Surveillance des animaux de rente : grâce à une géolocalisation en temps réel des vaches et de leurs mouvements auriculaires, le système IA auto-apprenant autrichien SmartBow permet de surveiller l'activité et la rumination de chaque animal. Les données sont immédiatement analysées en ligne, et l'agriculteur reçoit des informations en temps réel sous la forme de graphiques et d'alertes, qui lui signalent les comportements inhabituels, les dates de vêlage ou les périodes de chaleurs. En Suisse, ce système est utilisé par Agroscope.

Analyse prédictive : l'application de modèles de l'apprentissage automatique permet par exemple de suivre, voire de prévoir, les répercussions des facteurs environnementaux tels que les changements de conditions météorologiques sur le rendement des récoltes. Avec ce type de prédictions basées sur des modèles, l'agriculteur profite d'une rentabilité accrue de ses produits et d'une production plus économe en ressources. La start-up suisse Gamaya propose des solutions basées sur des données hyperspectrales et des analyses du *Big Data*.

Recherche sur les semences et l'amélioration des plantes : l'utilisation de l'IA permet une évaluation visuelle des germes et l'identification des semis par caractérisation phénotypique. Grâce à des outils d'analyse et à des algorithmes basés sur l'IA, il est ainsi possible de prédire les propriétés de nouvelles variétés à partir de l'ensemble du génome et de grandes séries de données phénotypiques. La société israélienne NRGene propose une plateforme *cloud* de ce type. En prenant en compte les effets d'interaction entre le génotype et les facteurs environnementaux, elle formule des recommandations sur les variétés suivant le site. Des fonctions similaires sont également disponibles pour l'élevage.

6.12.2 Défis

Outre les nombreuses applications qu'elle permet, l'utilisation de la technologie IA dans l'agriculture s'accompagne également de défis. Parmi eux, citons une infrastructure numérique insuffisamment développée localement, une base de données agronomiques encore souvent médiocres, une réceptivité hésitante de la part des agriculteurs et des coûts d'investissement parfois élevés pour les

utilisateurs potentiels. À cela s'ajoutent des incertitudes juridiques, par exemple concernant le traitement des données agricoles ou l'utilisation d'aéronefs et de véhicules autonomes (sans occupants).

Pour assurer une utilisation transparente et responsable de l'IA, il convient par ailleurs de tenir compte des réserves quant aux répercussions sur le changement structurel (avenir des petites exploitations agricoles, pertes d'emplois) et à la dépendance vis-à-vis des multinationales technologiques. D'un autre côté, l'IA a le potentiel de contribuer à l'accroissement de la compétitivité et de la durabilité de l'agriculture suisse, ainsi qu'à la simplification administrative et à la réalisation des objectifs de la politique agricole.

6.12.3 Activités existantes

L'Office fédéral de l'agriculture (OFAG) suit en permanence les développements de l'IA dans l'agriculture. À cet effet, il a créé le Business Intelligence Competence Center, opérant dans le domaine du *Big Data* et, ultérieurement, dans l'analyse prédictive. Pour encadrer l'utilisation des données et applications numériques a été lancée en 2018 la Charte sur la numérisation dans l'agriculture et le secteur agroalimentaire suisses, dont la mise en œuvre par la communauté réunie autour de la charte doit être encouragée. Comme le montrent les exemples précédents, la Suisse dispose d'un vivier dynamique de start-up. Le Conseil fédéral s'efforce d'améliorer en permanence les conditions-cadre applicables aux start-up, afin que notre pays reste attractif pour les jeunes pousses.

6.12.4 Évaluation et actions requises

Champ d'action 1 : Conséquences de l'IA sur l'agriculture	
Suivi des développements de l'IA dans l'agriculture	L'agriculture suisse devrait s'attaquer aux défis mentionnés. Dans ce contexte, l'OFAG suivra les développements de l'IA dans l'agriculture et traitera les questions soulevées dans le cadre des compétences existantes. **Responsabilité :** OFAG **Statut :** suivi dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.13 L'intelligence artificielle dans l'énergie, le climat et l'environnement

6.13.1 Vue d'ensemble

À partir de 2050, la Suisse devra cesser d'être un émetteur net de gaz à effet de serre. Pour y parvenir, il faudra que le secteur de l'énergie, entre autres, opère une transition au cours des 30 prochaines années afin de gagner en efficience et d'abandonner les combustibles fossiles. Parallèlement, il faudra mettre un frein à la consommation de ressources, actuellement trop élevée, et à l'appauvrissement de la biodiversité. La majeure partie des atteintes environnementales[108] est due à l'alimentation, au logement et aux transports – que ce soit en Suisse ou ailleurs. Transformer les systèmes pose de nombreux défis dans les domaines de l'approvisionnement en énergie, de l'alimentation, du logement et des transports. Or, l'IA est vue comme une technologie centrale pour répondre à ces défis.

Les technologies d'IA peuvent par exemple prédire la demande en énergie, en biens alimentaires et en autres biens de consommation de manière nettement plus précise qu'auparavant. Les secteurs productifs et commerciaux s'en servent déjà pour réduire la taille des stocks, les pertes liées à la planification de la production ou les erreurs en logistique des transports – y compris dans des contextes complexes, où la gestion du temps est un facteur critique ou qui font intervenir des chaînes de fournisseurs disséminés dans le monde. L'IA permet par ailleurs d'intégrer dans les processus de production la disponibilité des matières premières, l'état des écosystèmes sur les sites de production ou des aspects écologiques et sociaux plus généraux et de mettre ces informations à la disposition de la clientèle. Du côté de la demande, l'IA peut aider les consommateurs à trouver des variantes de produits qui répondent à leurs besoins individuels en tenant compte aussi de leurs critères environnementaux en plus des critères de prix.

Le **secteur de l'énergie** est un autre domaine dans lequel l'IA est amenée à jouer un rôle central. Né longtemps avant l'ère numérique, il est encore peu numérisé. Là aussi, l'IA peut être d'une grande utilité de plusieurs manières :

- en facilitant le passage d'un système très centralisé à un système décentralisé et modulable, qui ferait intervenir des énergies renouvelables et serait capable, au besoin, d'opérer un couplage plus étroit des différentes infrastructures énergétiques (électricité, gaz et chaleur) ;
- en optimisant la planification du réseau et les pronostics en matière de consommation et de production et en s'assurant d'une meilleure coordination dans un système toujours plus complexe, caractérisé par un approvisionnement fluctuant et décentralisé et par un nombre croissant d'acteurs ;
- enfin et surtout, en contribuant à réduire la consommation d'énergie et en soutenant les efforts visant à décarboner la consommation d'énergie tout en maintenant un niveau élevé de sécurité des approvisionnements et en limitant les coûts de la transition[109].

6.13.2 Défis

Les technologies d'IA sont un instrument à double tranchant dans le cadre des objectifs environnementaux. Actuellement, leur développement et leur utilisation nécessitent de grandes quantités d'énergie et de matières premières, et cette tendance devrait se poursuivre. À l'avènement de l'économie circulaire, la question de la compatibilité environnementale se posera avec une acuité croissante en ce qui concerne la fabrication, le recyclage et l'élimination des appareils et des infrastructures. Même si l'IA est potentiellement capable d'optimiser considérablement les processus, les produits et les marchés du point de vue écologique, ce potentiel n'est pas encore réalisé. Il est possible d'agir directement sur les technologies d'IA en développant des applications ciblées sur des objectifs environnementaux, ou indirectement, en mettant à disposition de grandes quantités de données environnementales pour développer des applications d'IA (approche incitative, ou « push »).

[108] Rapport du Conseil fédéral. « Environnement Suisse 2018 », 2018, disponible à l'adresse : https://www.bafu.admin.ch/bafu/fr/home/etat/publications-etat-de-l-environnement/environnement-suisse-2018.html

[109] Office fédéral de l'énergie. « La digitalisation du monde de l'énergie - Dialogpapier zum Transformationsprozess », 2019, disponible à l'adresse : https://www.bfe.admin.ch/bfe/fr/home/approvisionnement/digitalisation.html

Une telle approche a été adoptée pour le service européen de surveillance terrestre Copernicus mandaté par la Commission européenne.

Au sens d'une approche où le développement de l'IA est stimulé pour répondre à une problématique (approche « pull »), l'IA peut devenir une technologie clé, capable d'appréhender la complexité des systèmes évoqués pour atteindre les objectifs et les exigences en termes de consommation des ressources environnementales dans les domaines de l'approvisionnement en énergie, de l'alimentation, du logement et des transports. Cela exige toutefois que les objectifs et les exigences soient clairement définis, voire préexistants.

Dans le **secteur de l'énergie**, la situation est encore floue en ce qui concerne l'IA. La Stratégie énergétique 2050 pose un cadre important et contient des mécanismes « push » et « pull » visant la transition du secteur énergétique. On ne sait toutefois pas encore dans quelle mesure la numérisation du secteur encouragera le développement et l'utilisation de l'IA. L'apprentissage automatisé, par exemple, existe depuis assez longtemps mais il n'est guère utilisé. En lieu et place, ce sont plutôt des méthodes conventionnelles qui sont employées dans la planification et l'exploitation des infrastructures énergétiques. Il reste donc à analyser la manière dont l'IA pourrait être développée dans un secteur régulé, fragmenté et où les contraintes d'efficience sont encore peu marquées. Un marché de l'énergie ouvert présenterait des incitations supplémentaires vers des solutions et des prestations innovantes basées sur l'IA. Les approches en amont propres à favoriser le développement de l'IA et la numérisation sont encore peu établies dans la réglementation et les infrastructures qui offrent un access aux données digital ne sont pas suffisamment performantes pour autoriser l'usage de l'IA à grande échelle. En outre, il existe des incertitudes quant aux conséquences de l'IA sur la sécurité des approvisionnements. Il ne faut pas non plus sous-estimer la question épineuse de l'efficience énergétique. En conclusion, il reste encore à savoir comment l'IA pourrait soutenir l'efficience énergétique et les implications que cela pourrait avoir sur la protection des données. Donc, il faut travailler plus sur les bases necessaire et cadre de régles en supportant le dévélopment de l'IA.

6.13.3 Activités existantes

L'approche en amont, par l'encouragement des technologies d'IA dans le **domaine de l'environnement**, se concentre actuellement sur l'analyse de grandes quantités de données (telles que les données satellite). Les images satellite, par exemple, se sont multipliées ces dernières années. Depuis 2017, le service européen de surveillance terrestre Copernicus fournit plusieurs fois par semaine une image complète de la Suisse et plusieurs fois par mois une image complète de la surface terrestre. Sans le recours aux technologies d'IA, l'analyse de ces données serait impossible. Le champ des applications possibles est large et peut être de portée nationale (reconnaissance des récoltes dans les champs) ou internationale (détection d'incendies de forêt et d'opérations de déboisement). De même, depuis quelques années, les données tirées des réseaux nationaux de mesure environnementale sont rendues accessibles et alimentent de manière ciblée l'apprentissage des applications IA. Concernant les mesures prises en aval, la priorité est donnée aux objectifs à long terme de la politique environnementale. L'objectif de la neutralité carbone en Suisse d'ici à 2050 constitue en soi une forte incitation à développer les potentialités de l'IA. Les risques et les opportunités de cette approche « pull » en politique environnementale sont l'objet d'une étude mandatée par l'Office fédéral de l'environnement. Les conclusions sont attendues pour la fin 2019.

Dans le **secteur de l'énergie**, la stratégie énergétique 2050 recouvre des mesures plutôt en amont, sous forme d'objectifs. Elle pose les premiers jalons vers la création d'une infrastructure de données digital sans laquelle il serait impossible de faire intervenir l'IA. Ainsi, des systèmes de mesure intelligents (*smart metering*) seront introduits dans le domaine énergétique d'ici à la fin 2027. Cela permettra de numériser la saisie des données relatives à la production et à la consommation d'électricité, qui gagnera ainsi en précision. Dans un deuxième temps, il faudra discuter de la faisabilité d'une plate-forme numérique (Datahub) qui organisera les échanges de données plus efficacement et simplifiera la mise à disposition des données[110]. Les interfaces de programmation d'applications (API) auront un rôle important à jouer dans cet environnement. Une plate-forme associée à des API pourrait constituer le cœur de l'infrastructure de données évoquée. Il est par

[110] Office fédéral de l'énergie. « La digitalisation du monde de l'énergie - Dialogpapier zum Transformationsprozess », 2019.

ailleurs prévu de lancer le débat sur la numérisation du secteur de l'énergie afin de répondre aux questions sur les conditions-cadre du recours à l'IA. Les *regulatory sandboxes*, ces environnements de test sous contrôle étroit, seraient un outil possible pour tester de nouvelles conditions-cadre prévues spécifiquement pour les applications utilitaires de l'IA, dans la mesure où la réglementation actuelle empêche le développement de nouvelles applications dans certains domaines[111]. L'Office fédéral de l'énergie propose dans ce domaine des projets pilotes, des projets de démonstration et des projets phares qui offrent un cadre propice au lancement de projets d'IA[112].

6.13.4 Évaluation et actions requises

L'IA est un outil capable de contribuer à résoudre les contraintes écologiques posées aux systèmes d'approvisionnement en énergie, à l'alimentation, au logement et aux transports. Ses potentialités ont été jusqu'ici insuffisamment utilisées. Cet aspect est illustré par le fait que le secteur de l'énergie est peu représenté dans le portefeuille de brevets suisses (cf. chapitre 5, figure 11). Les besoins en savoir-faire et en capitaux pour mettre en place l'infrastructure correspondante aux applications IA (puissance de calcul, données) sont gigantesques. L'accès aux données environnementales et à une puissance de calcul suffisante pour faire fonctionner une application IA devrait être garanti.

L'IA appliquée au secteur de l'énergie et à l'environnement se trouve à un stade précoce et dynamique. En ce qui concerne l'énergie, les incitations et les infrastructures actuelles doivent être soumises à une réflexion critique et examinées sous l'angle de l'IA. Cet examen doit porter sur la structure du marché de l'énergie, la transparence, la performance de l'infrastructure de données et la régulation du réseau, laquelle tend à favoriser les investissements nécessitant beaucoup de capitaux. On en sait encore trop peu sur sur les champs d'application et l'utilité de l'IA, et le manque de personnel qualifié dans ce domaine est patent. Il faut donc d'abord analyser ces obstacles et améliorer les instruments nécessaires, par exemple en travaillant sur la disponibilité des données et la transparence. Il s'agira dans un deuxième temps d'approfondir les connaissances dans ce domaine, de les diffuser et de mieux informer sur ces thématiques. Dans ce contexte, il convient de réfléchir à la possibilité d'élaborer une stratégie sectorielle basée sur l'IA, assortie d'exemples concrets et de projets pilotes (par exemple pour identifier les obstacles posés par la réglementation).

Champ d'action 1 : Conséquences de l'IA sur le secteur énergétique L'IA a le potentiel d'améliorer considérablement l'efficience dans l'approvisionnement énergétique. Elle peut contribuer à développer les énergies renouvelables, à réaliser des économies d'énergie et, par conséquent, à soutenir la préservation du climat, une thématique qui gagne en complexité et qui est importante pour l'exploitation et la sécurité des approvisionnements.	
Suivi des développements dans le secteur énergétique	L'OFEN observera dans ce contexte les enjeux liés à l'utilisation de l'IA dans le secteur de l'approvisionnement en énergie, élaborera les bases nécessaires, développera les compétences et atténuera les barrières d'ordre réglementaire et technique. **Responsabilité :** OFEN **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

[111] L'Office fédéral de l'énergie a lancé une étude afin d'analyser les expériences réalisées au niveau international sur les *regulatory sandboxes* et leur application dans le secteur de l'énergie.

[112] Office fédéral de l'énergie. Programme pilote, de démonstration et programme phare, 2019, disponible à l'adresse : https://www.bfe.admin.ch/bfe/fr/home/recherche-et-cleantech/programme-pilote-de-demonstration-et-programme-phare.html

Champ d’action 2 : Conséquences de l’IA sur le climat et l’environnement L’IA est une technologie clé pour répondre aux défis environnementaux et systémiques que représentent l’alimentation, le logement et les transports. Il faudrait disposer des données nécessaires à cette fin (par exemple disponibilité des matières premières, état des écosystèmes sur les sites de production ou informations sur les processus de production) de manière simple et, dans la mesure du possible, les intégrer dans les flux d’information de la chaîne d’approvisionnement et des marchés. Les technologies d’IA vont très probablement occasionner une croissance supplémentaire de la consommation en ressources environnementales. Des questions liées à l’économie circulaire vont se poser avec davantage d’acuité, comme la compatibilité environnementale des produits, le recyclage et l’élimination des infrastructures et des appareils.	
Suivi des développements dans les domaines du climat et de l’environnement	L’OFEV veille à ce que les informations environnementales soient disponibles autant que possible sous la forme de jeux de données ouverts et présentés sous forme numérique, en faisant intervenir des applications IA. Il suit de près les difficultés pertinentes sur le plan environnemental en intégrant la perspective de l’économie circulaire et met à disposition les instruments nécessaires. **Responsabilité :** OFEV **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.14 L'intelligence artificielle dans l'administration

6.14.1 Vue d'ensemble

L'IA peut alléger l'administration à tous les niveaux, améliorer l'orientation vers la clientèle et la qualité de service et contribuer à l'accroissement de la rentabilité. L'utilisation de l'IA dans l'administration permettra de traiter rapidement, efficacement et 24 heures sur 24 les données qui ne pouvaient auparavant pas être traitées automatiquement. Des applications IA peuvent déjà être utilisées dans cinq domaines de l'administration : [1] reconnaissance de texte, [2] reconnaissance d'images et de vidéos, [3] aides à la traduction automatique (pour les documents administratifs), [4] analyse automatique d'enregistrements sonores, [5] interaction via des *chatbots*. Vous trouverez en annexe de ce rapport des informations complémentaires sur l'utilisation de l'IA dans l'administration.

6.14.2 Défis

Plusieurs conditions doivent être remplies pour assurer le succès du déploiement de l'IA dans l'administration. L'utilisation de systèmes d'intelligence artificielle requiert des volumes importants de données. Au sein de l'administration fédérale, on en trouve notamment à l'AFD (DaziT), l'AFC, l'OFS, l'OFAG, etc. C'est la raison pour laquelle ces unités administratives jouent un rôle central dans l'introduction de l'IA. Les échanges de données constituent un autre facteur de succès : ils doivent être réglés et simples entre les trois niveaux institutionnels et au sein des quelque 70 unités administratives de l'administration fédérale. En vertu de la protection des données, des limites sont fixées pour le flux de données à caractère personnel. Ce flux de données doit être amélioré pour que l'IA puisse déployer tout son potentiel.

L'IA permettra d'améliorer les prestations proposées à l'ensemble des clients de l'administration. Cependant, l'être humain devra rester – au moins à court ou moyen terme – la dernière instance de décision s'agissant des décisions administratives qui concernent le statut juridique du destinataire ou les prestations de l'État : la transparence, la traçabilité et le contrôle des résultats de l'administration doivent être préservés, ce qui n'est pas (encore) possible en l'état actuel de la technologie. L'IA est un domaine de recherche récent, avec peu de valeurs empiriques et des taux d'erreur pas toujours connus. Or, les activités de l'administration ne tolèrent qu'un très faible taux d'erreur.

Alors qu'il existe aujourd'hui une offre large de systèmes d'intelligence artificielle de base sous la forme de solutions propriétaires ou de solutions « open source », la modélisation et l'adaptation de ces systèmes constituent des défis majeurs. Ils nécessitent une expertise et une expérience qui ne sont disponibles que ponctuellement dans l'administration et qui sont en outre difficiles à maintenir. Une structure décentralisée de ces compétences et de leur gestion serait donc inefficace. Enfin, il convient de répondre aux doutes et à la peur de l'avenir des collaborateurs de l'administration en positionnant l'IA comme une chance et non comme un danger.

6.14.3 Activités existantes

À l'heure actuelle, l'administration fédérale ne dispose pas d'applications opérationnelles pouvant être déployées à grande échelle ou celles-ci en sont encore aux premiers stades de leur développement (étude ou prototype). Les projets suivants sont connus (état mars 2019) :

Tableau 6: Applications de l'IA dans l'administration fédérale

Projet	Administration	Description
Programme DaziT Projet Data Analytics	AFD	Le projet vise à ce que l'AFD soit parfaitement préparée à l'utilisation ciblée de l'analyse des données. Les cas d'application sont les suivants : profils d'entreprises, analyses des risques et contrebande de marchandises.
Programme DaziT Évaluation d'une solution de *chatbot*	AFD	Afin de réduire les dépenses liées aux passages de frontières (en particulier dans le domaine du personnel), l'objectif est d'atteindre un degré très élevé d'automatisation des processus. Une solution de *chatbot* est actuellement évaluée.

Projet	Administration	Description
Arealstatistik Deep Learning (ADELE)	OFS	Le projet vise à automatiser, au moins partiellement, l'interprétation des images aériennes à l'aide de l'IA en vue de la détection et de la classification des changements. Le *deep learning* est particulièrement adapté à ce projet, car de très grandes quantités de données d'entraînement sont disponibles.
Automatisation du codage NOGA (NOGAuto)	OFS	Le projet a pour but d'automatiser le codage des données qui sont déjà à disposition au sein de l'OFS (le codage est effectué manuellement à l'heure actuelle) à l'aide des méthodes de l'apprentissage automatique.
Apprentissage automatique SoSi	OFS	L'objectif du projet est de grouper les bénéficiaires de prestations en fonction de leur parcours dans le système de sécurité sociale et le monde du travail, et d'estimer la probabilité d'appartenance à un groupe donné au moyen d'une procédure d'apprentissage automatique. La fiabilité des estimations est mesurée.
Contrôles de plausibilité par des techniques du machine learning	OFS	Ce projet vise à développer et à rendre plus rapides les contrôles de plausibilité à l'OFS par l'utilisation d'algorithmes de machine learning. Il vise par là même à améliorer la qualité des données.
Essai pilote d'AA afin d'optimiser pour le marché de l'emploi la répartition de requérants d'asile dans un canton	SEM	Un algorithme basé sur l'AA est testé sur une période de temps et avec des quantités de données limitées afin d'optimiser la répartition des requérants d'asile dans un canton sous l'angle du marché du travail[113].

Source : UPIC.

Selon une enquête sur l'utilisation de l'intelligence artificielle par les cantons, ces derniers s'engagent dans la conception ou l'implémentation d'applications IA à tous les niveaux et dans divers domaines d'application[114]. Par exemple, la police de Lucerne utilise une solution IA d'analyse vidéo des infractions. Plusieurs cantons (dont les cantons de Saint-Gall, de Fribourg et de Lucerne) étudient des solutions de *chatbot* pour leurs services d'assistance aux utilisateurs. L'administration de la sécurité sociale du canton de Saint-Gall (SVA) a par exemple testé un *chatbot* de contact avec ses clients. Les habitants du canton peuvent se renseigner sur leur droit à la réduction des primes sur le *chatbot*.

Une comparaison internationale montre que beaucoup d'autres pays (notamment industrialisés) élaborent également des stratégies portant sur l'utilisation de l'IA dans l'administration. Certains sont nettement plus avancés dans la formulation et la mise en œuvre concrète. Si la Suisse ne répond pas de manière intensive et coordonnée aux opportunités et aux risques de l'IA, l'écart continuera de se creuser[115].

[113] Cf. EPF Zurich. « Algorithmus verbessert Erwerbschancen von Flüchtlingen », 2018, disponible à l'adresse https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2018/01/algorithmus-verbessert-erwerbschancen-von-fluechtlingen.html
et Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence. «Improving refugee integration through data-driven algorithmic assignment», *Science*, Vol. 359, Issue 6373, 2018, pp 325-329, https://science.sciencemag.org/content/359/6373/325

[114] Cf. résultats de l'enquête sur l'utilisation de l'intelligence artificielle (IA) par les cantons du 15 juillet 2019, canton de Lucerne, manuscrit non publié.

[115] Au niveau international, des pays comme la Chine, la Finlande ou Singapour considèrent que l'IA est une technologie clé de la numérisation. Ils ont annoncé des plans visant à promouvoir massivement et rapidement la recherche et le développement en la matière. La Commission européenne promeut, elle aussi, l'IA et a formulé trois lignes d'action : [1] encouragement à l'investissement, [2] préparation aux changements socio-économiques et [3] établissement d'un cadre éthique et juridique approprié.

6.14.4 Évaluation et actions requises

Il est trop tôt pour des considérations concrètes et des recommandations concernant les processus de décision algorithmiques ou automatiques en découlant. Il ne faut confier aucun pouvoir de décision autonome aux systèmes basés sur l'IA, mais les définir comme des fonctions de soutien permettant aux collaborateurs de l'administration de prendre de meilleures décisions plus rapidement. Les recommandations seront formulées à la suite des expériences qui ressortiront des premiers projets concrets et serviront de garde-fous à les projets suivants. Il ressort de l'analyse les champs d'action suivants :

Champ d'action 1 : Bases de données communes Des bases de données communes au sein de l'administration offrent plus de possibilités d'application pour l'IA, car davantage de données sont disponibles, et des relations plus fortes peuvent être établies entre elles.	
Création et mise à disposition de bases de données au sein de l'administration fédérale	Le développement de l'IA nécessite des bases de données volumineuses et protégées. L'administration fédérale doit créer des bases de données et les mettre à disposition afin de permettre des échanges productifs de données. Des prescriptions garantissant une utilisation sécurisée des données doivent être définies. **Responsabilité :** Offices fédéraux traitant de grandes quantités de données (notamment l'AFD, l'OFS, l'AFC et l'OFAG) **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d'action 2 : Réseau de compétences IA au sein de l'administration fédérale L'identification de processus dans l'ensemble de l'administration et l'accès transversal aux données constituent des conditions indispensables pour exploiter le potentiel de l'IA dans l'administration fédérale.	
Clarifications approfondies concernant la création d'un réseau de compétences IA spécifiquement consacré aux aspects techniques de l'utilisation de l'IA dans l'administration fédérale	L'identification de processus dans l'ensemble de l'administration et l'accès transversal aux données constituent des conditions indispensables pour exploiter le potentiel de l'IA dans l'administration fédérale. Le développement et l'échange de connaissances et d'expériences au niveau interdépartemental sont essentiels pour un développement économique et coordonné de solutions IA dans l'administration fédérale. Les solutions fragmentées ne sont en revanche pas efficaces. Un point de contact unique ou un réseau de compétences axé sur les aspects techniques de l'application concrète de l'IA au sein de l'administration fédérale pourrait être une solution. Ce point de contact ou ce réseau devrait notamment remplir un rôle consultatif. Le DFF (UPIC), en collaboration avec le DFI (OFS) et avec la participation des autres départements et de la ChF, étudie la valeur ajoutée (avec notamment une analyse des besoins) et la faisabilité d'un point de contact ou d'un réseau de compétences, en portant une attention particulière aux aspects techniques de l'application de l'IA dans l'administration fédérale. Cet organe doit avoir un rôle consultatif pour ce qui concerne l'application de l'IA dans l'administration fédérale. En vue d'une mise en réseau complète et d'une réflexion technologique globale, il faudrait en outre prendre en compte d'autres technologies de la transformation numérique (p. ex. la blockchain ou l'IdO, etc.) et les défis liés à leur mise en œuvre **Responsabilité :** UPIC, OFS **Statut :** examen à l'attention du Conseil fédéral
Actions supplémentaires requises : oui	

Champ d’action 3 : Mettre en avant les opportunités offertes par l’IA (communication) Une information active et la présentation des opportunités offertes par l’IA peuvent atténuer les craintes des collaborateurs.	
Renforcement de la communication sur les thématiques liées à l’IA au sein de l’administration fédérale	L’IA modifiera également certains profils de poste dans l’administration fédérale et augmentera sensiblement la pression à l’adaptation. Il faut accompagner le processus de changement afin que les collaborateurs le vivent de manière positive et le soutiennent activement. **Responsabilité :** UPIC **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d’action 4 : Bases légales, contrôle et sécurité des données L’IA doit pouvoir être déployée au sein de l’administration fédérale. Les obstacles inutiles doivent être éliminés. Il faut créer un environnement empreint d’esprit pionnier.	
Examen des bases légales en vue de l’utilisation de l’IA dans l’administration fédérale	L’utilisation de l’IA nécessitera une adaptation de la législation à moyen ou long terme. Il convient notamment de clarifier les conditions-cadre en matière de contrôle des données (cybercriminalité, garantie de la sécurité des données). Dans ce contexte, la prochaine révision de la LPD est prise en compte. **Responsabilité :** UPIC en étroite collaboration avec l’OFS et les offices fédéraux concernés **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.15 Développement du cadre juridique général au regard de l’intelligence artificielle

6.15.1 Vue d’ensemble

Comme nous l’avons vu au chapitre 4, le cadre juridique général est (1) parfaitement applicable à l’utilisation de l’IA, (2) selon l’analyse de l’état actuel de la technologie présentée dans le présent rapport, globalement adapté aux nouveaux modèles de gestion et applications, et (3) une éventuelle adaptation des normes juridiques n’est pour l’heure indiquée que pour des applications spécifiques.

Comme expliqué au chapitre 3, le spectre des applications IA actuelles et probables reste en outre très limité. Les développements progressent cependant à un rythme fulgurant. Aussi, des questions juridiques, éthiques et sociales soulevées par les possibilités de l’IA se poseront de manière aiguë dans les années à venir.

6.15.2 Défis

Développement du cadre juridique
Le développement de nouvelles applications est susceptible de poser des défis majeurs, y compris dans le cadre juridique général en vigueur. À la problématique exposée au chapitre 4 sur l’explicabilité, la transparence et le risque de discriminations s’ajoute celle de la capacité des systèmes d’IA à agir de manière de plus en plus autonome, ce qui met à l’épreuve le cadre juridique actuel. L’intelligence artificielle peut potentiellement remettre en cause des prémisses essentielles de l’État de droit, qui repose sur le principe selon lequel l’homme décide et agit et prévoit des procédures de décision reposant sur l’humain (p. ex. législation, jurisprudence, élections, votations, etc.).

Le recours à l’IA dans le secteur public, que ce soit dans l’administration ou dans le domaine judiciaire, devrait être systématiquement évalué sur la base d’une liste de critères assurant la fiabilité du processus, sa transparence, sa neutralité et son intégrité, le respect des droits de l’homme, notamment des droits fondamentaux et l’absence de discrimination, de même que l’absence de caractère prescriptif. Mais le risque d’atteinte à ces droits ne se limite évidemment pas au secteur public. Le Conseil fédéral suit attentivement cette évolution. S'il devait constater que les bases légales actuelles ne répondent pas de manière satisfaisante à ce risque, il prendrait les mesures nécessaires, notamment en s'inspirant des solutions du type de celles instituées pour d’autres domaines (p. ex. recherche sur l’être humain ou la loi sur l’analyse génétique humaine).

Dans un tel contexte, il faudrait notamment vérifier si, du point de vue de l’État de droit, des dispositions doivent être prises pour déterminer comment la transparence de ces systèmes peut être favorisée et comment expliquer la logique sur laquelle ces systèmes reposent leurs décisions sous une forme compréhensible pour l’homme de sorte que l’explicabilité et, le cas échéant, la contrôlabilité des systèmes puissent être garanties. Les secrets d’affaires et de fabrication doivent être pris en compte (voir aussi mesures envisagées dans le cadre de la révision LPD, ch. 4.3).

La transparence peut aussi concerner l’utilisation de l’IA dans les interactions avec la clientèle. Il existe déjà des systèmes capables de mener un entretien téléphonique avec des êtres humains sans que ces derniers s’en rendent compte (cf. chapitres 3.3 et 4.3). Se pose dès lors la question de savoir si les entreprises qui en font usage doivent en informer leurs clients au préalable.

Les systèmes d’IA ont besoin de très grandes quantités de données. Plus particulièrement, dans les cas où les pouvoirs publics utilisent des applications IA, il faut garantir que les éventuels problèmes de qualité des données n’entraînent pas des biais, des inégalités de traitement inacceptables, des discriminations ou tout autre préjudice pour les personnes concernées.

Dans certains cas, les questions soulevées par les systèmes d’IA peuvent dépasser les frontières. Mais en dépit du nombre croissant d’affaires internationales, il ne semble pas, de prime abord, y avoir de nouveaux problèmes qui ne puissent être résolus par le droit international privé en vigueur[116]. En

[116] Dans le cas du fondement de la responsabilité extracontractuelle, la compétence se situe par exemple en Suisse en tant que lieu du résultat (art. 129 LDIP ; art. 5, ch. 3, CL), et le droit suisse serait régulièrement d’application (pour la responsabilité extracontractuelle, cf. p. ex. art. 133 LDIP ; pour les défauts de produits, cf. art. 135 LDIP ; pour les accidents de la route, cf. art. 3 de la Convention de La Haye de 1971 [RS 0.741.31]).

droit pénal également, le contexte international ne crée a priori pas de nouveau problème qui soit insoluble au moyen des instruments existants (notamment les art. 3 ss CP et les conventions internationales d'entraide).

Il faut toutefois rappeler que l'entraide judiciaire en matière pénale dans le domaine de la cybercriminalité est liée à des difficultés additionnelles. Ces difficultés peuvent survenir également dans l'entraide judiciaire pour des crimes commis au moyen de systèmes d'IA ou qui touchent des systèmes d'IA.

Futurs développements de l'intelligence artificielle et défis éthiques

Si le présent rapport est consacré aux applications actuellement rendues possibles par l'IA, on ne peut pas exclure que l'évolution des technologies s'accompagne de défis juridiques, éthiques et sociaux d'un nouveau genre. Il convient donc de poursuivre le dialogue sur les questions générales soulevées par l'IA et qui remettent en question, voire menacent notre système de valeurs actuel.

Dans certains domaines spécifiques où il n'est pas possible de développer de nouvelles applications en raison de la réglementation en vigueur, on peut se tourner vers le *sandboxing*. Dans cet environnement de test contrôlé et sécurisé, qui doit être le plus flexible et donc léger sur le plan administratif, l'assouplissement de certains paramètres (p. ex. octroi facilité des autorisations à titre d'essai et simplification des procédures applicables) permet d'encourager le développement de nouvelles technologies et modèles de gestion innovants, tout en testant les exigences légales minimales requises ou les points sur lesquels le droit doit faire preuve d'une certaine flexibilité. Les connaissances et la transparence apportées par une *sandbox* aident à adapter le cadre réglementaire en temps utile, permettant à l'économie de mieux tirer parti de l'innovation.

Il convient également de tenir compte des activités menées au plan international. Le processus d'élaboration de normes mené par des sociétés technologiques transnationales puissantes, de même que les principes éthiques élaborés par diverses instances internationales, souffrent d'un déficit démocratique. Pour la Suisse, la question est de savoir comment prendre part à ces processus de manière pertinente et s'y positionner de sorte qu'elle puisse non seulement exprimer ses valeurs et conceptions en la matière, mais aussi évaluer les répercussions possibles sur l'ordre juridique national.

6.15.3 Activités existantes

Cadre juridique en vigueur

Plusieurs clarifications ont montré que la réglementation existante en matière de responsabilité civile et pénale et de droit international privé était à ce jour suffisante pour ce qui concerne la question de la responsabilité de machines agissant de manière autonome. Jusqu'à présent, il n'est pas apparu que son application aux robots crée des lacunes en matière de responsabilité. Aucune nécessité de mesures législatives n'a encore été identifiée[117]. La question de la transparence des applications d'IA a été traitée lors de la révision de la loi sur la protection des données dans la mesure où elle concerne des décisions automatisées uniques.

Futurs développements de l'intelligence artificielle et défis éthiques mondiaux

Étant donné la dimension internationale du développement de l'IA, il est important que la Suisse suive de près les efforts de réglementation au plan international. En particulier, elle doit s'attacher à comprendre comment un nouveau droit international est formé dans ce domaine et quelles peuvent en être les répercussions pour notre pays. La Confédération est déjà activement impliquée dans ces débats internationaux, en œuvrant notamment pour que les valeurs et normes établies soient respectées et toutes les parties prenantes impliquées.

6.15.4 Évaluation et actions requises

Le présent rapport identifie les domaines thématiques qui dans lesquels l'IA soulève des questions d'ordre juridique qui requièrent une clarification ou, a minima, une observation étroite. Aucune action n'est actuellement requise concernant le cadre juridique national. Le contexte international, par contre,

[117] Cf. p. ex. 18.3445 Ip. Marchand-Balet concernant les véhicules autonomes ; 17.3040 Po. Reynard « Examen de la création d'une personnalité juridique pour les robots » et rapport du Conseil fédéral en réponse au postulat Leutenegger Oberholzer 14.4169 « Automobilité » concernant les faits ayant un lien avec l'étranger.

appelle des mesures. Il est en effet prioritaire de produire et de mettre en œuvre des normes de droit international.

Champ d'action 1 : Formation d'un droit international spécifique de l'IA Il faut examiner de façon plus approfondie la manière dont les règles internationales relatives à l'IA sont définies et qualifiées, dans quelle mesure elles produisent un droit international et quelles pourraient en être les répercussions pour la Suisse.	
Rédaction d'un rapport sur l'évolution du droit international dans le domaine de l'IA	Un rapport sur l'évolution du droit international dans le domaine de l'IA doit être remis au Conseil fédéral d'ici fin 2020. Le cas échéant, le rapport pourra proposer des mesures permettant à la Suisse de se positionner sur la question. **Responsabilité :** DFAE **Statut :** examen à l'attention du Conseil fédéral
Actions supplémentaires requises : oui	

Champ d'action 2 : **Identification des systèmes d'IA dans les interactions avec les consommateurs**	
Suivi de l'évolution dans les interactions avec des systèmes d'IA	En principe, la loi fédérale contre la concurrence déloyale (LCD) est une base légale appropriée dans ce qui touche à l'identification de l'IA dans ses interactions avec des clients (chatbots, appels téléphoniques menés par des IA). Il faudrait toutefois qu'une plainte concrète soit déposée pour pouvoir approfondir la question. **Responsabilité :** DEFR (SECO) **Statut :** observation dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.16 Utilisation de l'intelligence artificielle dans la justice

6.16.1 Vue d'ensemble

Le potentiel de l'IA en tant qu'outil du système judiciaire est considérable. Certains outils développés aujourd'hui visent à aider les professionnels du droit à effectuer des recherches juridiques ou à anticiper l'issue possible d'une affaire portée devant un tribunal (les instruments dits de « justice prédictive »). D'autres peuvent être utilisés pour aider les tribunaux dans la gestion des affaires (par exemple en examinant et en attribuant les demandes aux sections judiciaires responsables) ou pour analyser le rendement des tribunaux. De plus, ces outils peuvent être utilisés pour faciliter la résolution des litiges en ligne. En Suisse, l'IA est actuellement peu utilisée pour prédire les décisions de justice, mais le sera beaucoup plus dans un avenir proche. Dans ce cadre, les décisions sont simulées au moyen de données historiques.

Toutefois, les décisions rendues par la justice sont particulièrement sensibles compte tenu des droits des personnes concernées. Il faut donc être particulièrement attentif aux problèmes de transparence et de discrimination que peut poser l'IA. Dans cette mesure, l'intégration de systèmes d'IA dans la structure de l'État de droit impose que les décisions (rendues avec ou sans aide humaine) puissent éclairer sur leurs motifs sous une forme compréhensible pour l'être humain.

6.16.2 Défis

Le recours à des décisions automatisées dans le domaine étatique (décisions administratives et judiciaires) peut faire sens dans certains domaines, mais ne devrait pas limiter le droit d'être entendu ou le droit à une décision motivée et sujette à recours. Il conviendra de veiller au respect de cette exigence dans les cas où l'IA présente un risque élevé sur le plan des droits de la personne concernée. Il convient d'éviter que la prise de décision soit en quelque sorte déléguée à une machine peu fiable[118]. Le Conseil de l'Europe a élaboré une charte éthique sur l'utilisation de l'IA dans les systèmes judiciaires et leur environnement[119] qui définit des principes à respecter et délimite en différentes catégories de risques les différentes utilisations possibles.

Du fait du traitement de données orientées vers le passé, l'analyse prédictive (*predictive analytics*) [120] renferme le risque de maintenir le statu quo dans la jurisprudence, laissant peu de place au développement du droit. Les circonstances extérieures, les événements irrationnels/émotifs et les exceptions sont tout aussi peu pris en compte. Le recours à l'analyse prédictive pour contrôler les juges est délicat en cela que ces derniers devraient rendre compte de leurs jugements à une technologie non exacte. Du point de vue des droits fondamentaux et des droits de l'homme, ce risque pourrait entraîner le non-respect du droit à bénéficier d'un tribunal indépendant et impartial (art. 30 Cst). En effet, le tribunal pourrait subir une pression extérieure en raison de l'utilisation de l'IA. Les prédictions automatisées des chances de réussite sont perçues comme étant objectives et exercent en conséquence une forte influence sur les décisions des parties, voire du tribunal dans certains cas. Le recours à des algorithmes de prédictibilité pourrait conduire à des refus plus nombreux (ex : refus si moins de 50% de chances de succès vs. refus si les chances de succès paraissent manifestement moins importantes que les risques d'échec). Du point de vue des droits fondamentaux et des droits de l'homme, ce risque peut déboucher sur une atteinte au droit d'accès à un tribunal (art. 29a Cst, art. 6, ch. 1, CEDH et art. 14, al. 1, Pacte II de l'ONU).

---

[118] Le rapport du Conseil européen (voir référence ci-dessous) classe les opportunités et les risques en quatre catégories : utilisations à encourager, utilisations à envisager avec de fortes précautions méthodologiques, utilisations à envisager au terme de travaux scientifiques complémentaires et utilisations à envisager avec les plus extrêmes réserves. L'analyse prédictive entre dans la deuxième catégorie et l'utilisation aux fins de contrôle des juges dans la troisième.

[119] Réf. Commission européenne pour l'efficacité de la justice. « Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement », 2018, https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b

[120] L'analyse prédictive a le potentiel de bouleverser le conseil juridique et la justice. Exemples : les honoraires des avocats peuvent être calculés en fonction de la difficulté du dossier (prédictions relatives aux chances de réussite) et les « performances » des avocats peuvent être comparées ; ajustement des primes d'assurance selon le risque de procès ; les sociétés de financement de procès peuvent choisir leurs dossiers plus efficacement ; les prédispositions des juges peuvent être établies par l'IA.

En conséquence, l'État doit veiller à ce que l'accès au droit ne soit pas dénié pour des raisons économiques ni pour des raisons liées à l'IA. Cela concerne plus particulièrement les parties qui ne sont pas sans ressources. Mais le calcul de faibles chances de réussite par l'IA peut aussi poser problème sur le plan des droits fondamentaux et des droits de l'homme pour les parties sans ressources, car ces dernières pourraient se voir refuser l'assistance judiciaire au motif que le recours apparaît d'emblée voué à l'échec sur la base de ce calcul.

L'intelligence artificielle peut aussi être utilisée pour rendre des décisions, pour détecter des incohérences ou comme soutien à la décision (par ex. calcul des dommages-intérêts). Il est important que le juge puisse s'écarter des conclusions de la machine et/ou que la décision puisse être revue par une personne physique.

6.16.3 Activités existantes

Le projet de révision LPD concerne également l'emploi de l'IA dans la justice ; il prévoit l'exigence d'une base légale formelle lorsque le traitement de données constitue un profilage ou que la finalité ou le mode de traitement est susceptible de porter gravement atteinte aux droits fondamentaux de la personne concernée.

Au cours des prochaines années, les autorités de poursuite pénale et les tribunaux de la Confédération et des cantons vont introduire conjointement sur l'ensemble du territoire la communication électronique ainsi que la gestion électronique des documents et des dossiers reconnue sur le plan juridique. Avec le projet « Justitia 4.0 », l'ensemble des documents, notes et dossiers d'une procédure judiciaire doivent pouvoir être enregistrés et échangés par voie électronique. Les parties (juges, avocats, autorités, etc.) peuvent consulter les entrées les concernant sur un portail centralisé. Cependant, aucune utilisation concrète de l'IA n'est prévue à l'heure actuelle.

6.16.4 Évaluation et actions requises

Champ d'action 1 : **Aide à la décision basée sur l'IA dans l'administration et la justice (analyse prédictive)**	
Observation de l'évolution de l'aide à la décision basée sur l'IA dans l'administration et la justice (analyse prédictive)	L'utilisation d'applications basées sur l'IA pour élaborer des bases de décision, préparer des décisions ou aider les décideurs peut être intéressante aussi bien dans l'administration que dans le système judiciaire. Cependant, comme le montrent les expériences en la matière menées dans d'autres pays et à l'échelle internationale, cela comporte aussi des risques. En cas d'utilisation de ce type de technologies dans l'administration ou la justice, l'autorité administrative ou judiciaire compétente doit donc vérifier au cas par cas si ces technologies ont des répercussions particulières sur les droits fondamentaux et procéduraux des parties et si ces répercussions nécessitent des bases légales spécifiques. Dans ce contexte, l'autorité administrative ou judiciaire compétente peut se référer aux travaux d'organisations intergouvernementales (p. ex. *Ethics guidelines for trustworthy AI* ou chapitre IV des recommandations de l'OCDE. **Responsabilité :** OFJ, DFAE **Statut :** suivi dans le cadre des compétences existantes
Actions supplémentaires requises : non	

6.17 Intelligence artificielle, données et droit de la propriété intellectuelle

6.17.1 Vue d'ensemble

La disponibilité des données joue un rôle essentiel pour l'application des méthodes IA actuelles. Pour la Confédération, il en découle des questions importantes concernant l'élaboration de la **politique des données**. Celle-ci vise en premier lieu à favoriser l'accès aux données, notamment à des données librement accessibles (*Open Data*) en tant que matière première d'une économie et d'une société numériques, ainsi qu'à instaurer des bases légales et des conditions-cadre cohérentes et adaptée aux réalités actuelles, permettant à la Suisse de se positionner comme un pôle attractif en matière de création de valeur par les données. Mais la politique des données définit également le cadre juridique à l'intérieur duquel les données peuvent être recueillies, liées et analysées de manière licite. Il existe des conflits d'objectifs avec les exigences de la **protection des données**, qui doit s'adapter aux nouvelles possibilités d'analyse de très grandes quantités de données offertes par l'IA. La propriété intellectuelle constitue une incitation à la création et à l'innovation tout autant qu'un outil de diffusion des connaissances.
La production de biens immatériels pourrait être facilitée et augmenter significativement avec l'aide de l'IA dans les activités de création et de R-D. À terme, l'IA pourrait prendre une part toujours plus importante au sein des procédés créatifs et inventifs. Ce constat devrait conduire à une évolution du droit de la propriété intellectuelle. De plus, parmi les données dont se nourrit l'IA se trouvent des données **protégées par un droit de propriété**. Il peut s'agir par exemple d'un droit d'auteur pour un texte, un dessin ou une photographie.

6.17.2 Défis

La plupart des applications IA ont besoin d'importants volumes de données sous une forme adéquate. La principale question qui se pose à la Confédération est de savoir comment encourager la disponibilité des données. Dans le domaine de la **politique des données**, l'absence de standard commun est l'un des obstacles le plus fréquemment cité. Par ailleurs, les données doivent être trouvables, c'est-à-dire qu'elles doivent être classées et/ou pouvoir être recherchées et réutilisées.

Les analyses basées sur l'IA permettent aussi de nouvelles formes d'analyse, qui confrontent la **protection des données** à des défis inédits. La capacité croissante de l'IA à mettre en relation des ensembles de données et à comparer différents types d'informations rend de plus en plus difficile la distinction entre données personnelles et données non personnelles. De plus, les systèmes d'IA ont la possibilité de déduire des informations personnelles à partir (d'une combinaison) d'éléments de données non personnelles. Ces données qui peuvent être identifiées par l'IA alors qu'elles n'ont initialement pas de caractère personnel soulèvent des questions en matière de consentement, de finalité et d'utilisation.

En outre, le traitement des données est de moins en moins transparents ; par conséquent, il devient difficile de fournir les informations appropriées aux personnes concernées et de cerner le rôle joué par les données dans la prise de décision. Il existe aussi un risque que les algorithmes renforcent les discriminations (même de façon non intentionnelle) dans les cas où ils fondent leurs résultats sur des données de mauvaise qualité. La marge de manœuvre de la Suisse dans le domaine de la protection des données est faible. S'éloigner du standard fixé par l'UE pourrait avoir des conséquences négatives en entravant la libre circulation des données.

Lors du traitement d'une donnée par l'IA, une copie numérique de la donnée est réalisée. Or, si la donnée traitée est protégée, par exemple par un **droit d'auteur**, cette copie numérique constitue une violation du droit d'auteur. Il est légitime de se demander à quelles conditions ce type de données peut être utilisé par des tiers, puisqu'il en va de la garantie constitutionnelle de la propriété pour les titulaires ou les ayants droit. Si, à l'avenir, l'IA permet de produire toujours plus de biens immatériels et à un coût toujours plus faible, le rôle et la place de la PI pourraient être appelés à évoluer. Lorsque les systèmes d'IA auront acquis la capacité à effectuer des choix indépendants et seront ainsi devenus inventifs ou créatifs se posera la question de la protection conférée à ces productions par la PI. S'il n'est plus possible d'attribuer avec certitude un acte créateur ou inventif à un être humain ou à un système d'IA, les critères de « création de l'esprit » et « d'activité inventive », intrinsèquement liés à la

nature humaine, ne seront plus véritablement utilisables. Il se pourrait alors qu'un changement de système soit nécessaire.

6.17.3 Activités existantes

Des progrès ont été réalisés au cours des dernières années dans le domaine du libre accès aux bases de données (*Open Data* ou *Open Government Data, OGD*), que ce soit dans l'administration fédérale ou les entreprises liées à la Confédération. À travers la **stratégie OGD**, le Conseil fédéral promeut la disponibilité, la transparence et l'efficacité de l'utilisation des données publiques.

Dans le domaine de la protection des données personnelles, le **projet de révision de la LPD**, en cours d'examen par le Parlement, tient déjà compte de certains enjeux liés à l'IA. Plusieurs mesures prévues revêtent une importance particulière pour la protection de la sphère privée dans le domaine de l'IA (cf. chapitre 4)[121]. En outre, la question des bases légales et de leur actualité en lien avec les données personnelles et non personnelles a été examinée dans le cadre de la Stratégie Suisse numérique en 2017. Cet examen n'a pas révélé de besoin fondamental de revoir les bases légales actuelles au-delà du projet de révision LPD et de besoins d'agir ponctuels. Le groupe d'experts[122] sur le traitement et la sécurité des données s'est aussi penché de manière extensive sur les questions de traitement et d'accès aux données personnelles et non personnelles en lien avec la numérisation, y compris en lien avec l'IA. Ces travaux ont débouché sur des recommandations qui sont en cours d'examen.

Le **projet de révision de la loi fédérale sur le droit d'auteur et les droits voisins** contient une disposition qui restreint le droit d'auteur et autorise expressément le « text and data mining » dans le cadre de la recherche fondamentale. Cela facilitera considérablement la recherche et contribuera à renforcer le pôle de recherche suisse.

6.17.4 Évaluation et actions requises

Champ d'action 1 : Politique des données	
Poursuite des travaux en cours sur la politique des données de la Confédération	La disponibilité et l'accès des données sont essentiels. À cet effet, la Confédération conduit une politique des données diversifiée. Avec la stratégie OGD, le Conseil fédéral a renforcé la disponibilité et l'accès des données publiques. Les travaux ont déjà été lancés et peuvent être menés dans le cadre des activités existantes de l'administration. **Responsabilité :** OFCOM, SG DFI/OFS **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

[121] Tel est en particulier le cas de l'obligation de procéder à une analyse d'impact, des obligations en matière de transparence des responsables de traitement et des règles applicables aux décisions automatisées, du principe de protection des données dès la conception et par défaut, de l'obligation de signaler les violations de sécurité, de l'incitation à adopter des codes de conduite.

[122] Rapport du groupe d'experts concernant le traitement et la sécurité des données, 2018, disponible à l'adresse https://www.newsd.admin.ch/newsd/message/attachments/55754.pdf

Champ d’action 2 : Protection des données	
Poursuite des travaux en cours sur la protection des données	Pour ce qui concerne la protection des données, il se pose la question de savoir si des mesures ne devraient pas intervenir de manière sectorielle et non par le biais d’une refonte des bases légales générales. Se pose également la question de savoir si la solution aux problèmes constatés ne doit pas être, au moins en partie, technique plutôt que seulement juridique. De nombreuses mesures sont en cours d’examen (révision LPD en débat au Parlement, mandat d’examen du Conseil fédéral concernant la portabilité, recommandations du groupe d’experts sur le traitement et la sécurité des données) et la Suisse peut difficilement agir de manière isolée dans ce domaine. Par conséquent, d’abord de mener à terme les travaux en cours et d’observer parallèlement l’évolution de la situation tant au plan technique qu’international (en particulier au niveau de l’UE et du Conseil de l’Europe). **Responsabilité :** OFJ **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Champ d’action 3 : Propriété intellectuelle	
Poursuite des travaux en cours sur la propriété intellectuelle	Il n’y a pas de nécessité urgente de modifier le droit de la propriété intellectuelle. Dans l’immédiat, la législation en la matière jouit de suffisamment de flexibilité. Les autorités administratives ou judiciaires disposent d’une marge de manœuvre pour faire évoluer les fictions juridiques/notions juridiques indéterminées et ainsi accompagner l’émergence de l’intelligence artificielle. Cependant des discussions et des réflexions doivent être menées dès à présent afin de s’assurer que la PI demeure un outil utile au développement de l’IA à moyen et long terme. **Responsabilité :** IPI **Statut :** mise en œuvre dans le cadre des compétences existantes
Actions supplémentaires requises : non	

Annexe 1 : Vue d’ensemble des champs d’action

Ce rapport présente un nombre important de mesures, d’initiatives et de clarifications antérieures dans les domaines thématiques étudiés, qui sont traitées en premier lieu dans le cadre d’activités, de compétences et de procédures établies. Les activités ci-dessous sont essentielles à la poursuite des nombreux travaux de la Confédération en lien avec les défis relatifs à l’intelligence artificielle:

Organes internationaux et IA (OFCOM, DFAE)

Champ d’action 1 : échange d’informations et de connaissances et coordination des positions défendues par des représentants de la Confédération au sein d’organes internationaux

- Utilisation de la « plateforme tripartite » comme réseau de compétences national interdisciplinaire et pour la coordination des postes occupés par des représentants de la Confédération au sein d’organes internationaux dans le domaine de l’IA.

Champ d’action 2 : gouvernance globale

- Renforcement de la gouvernance globale
- Prise en compte de la problématique de l’IA dans la stratégie de politique étrangère 2020 - 2023

Champ d’action 3 : la Genève internationale

- Amélioration de l’interconnexion et de la collaboration entre les acteurs du domaine de l’IA
- Étude des possibilités de renforcement de la collaboration en vue de l’«AI for Good Summit»
- Renforcement de la « Geneva Internet Platform »
- Renforcement de la Genève internationale en tant que place forte de la gouvernance numérique dans la stratégie de politique étrangère 2020 – 2023

Programme « Europe numérique » (SEFRI et autres)

Champ d’action 1 : participation suisse aux programmes « Horizon Europe » et « Europe numérique »

- Examen de la participation suisse aux programmes « Horizon Europe » et « Europe numérique »

Changements dans le monde du travail (SECO)

Champ d’action 1 : conséquences de l’IA sur le marché du travail

- Suivi de l’évolution du marché suisse du travail

L’IA dans l’industrie et les services (SECO)

Champ d’action 1 : l’IA dans l’économie

- Suivi de l’évolution de l’« industrie 4.0 »

L’IA dans la formation (SEFRI, cantons et autres acteurs concernés)

Champ d’action 1 : assurer la transmission des compétences adéquates

- Assurer la transmission des compétences nécessaires à l’utilisation de l’intelligence artificielle à tous les niveaux de formation

Champ d’action 2 : assurer une utilisation transparente et responsable de l’IA dans la formation

- Assurer une utilisation transparente et responsable de l’IA dans la formation

Utilisation de l’IA dans le domaine des sciences et de la recherche (SEFRI)

Champ d’action 1 : compétences dans le domaine de la recherche et de l’innovation

- Assurer les compétences de recherche et de TST dans le cadre de la politique en matière de formation, de recherche et d’innovation (FRI)

L'IA dans le domaine de la cybersécurité et de la politique de défense (DFAE, armasuisse (DDPS), SRC, Centre de compétences pour la cybersécurité (DFF), ChF, Armée (SG-DDPS), OFPP, DEFR, DDPS)

Champ d'action 1 : conséquences sur la politique de sécurité étrangère

- Examen des implications liés à l'utilisation des systèmes basés sur l'IA pour la politique étrangère

Champ d'action 2 : formes de menace et doctrine

- Examen de la cybersécurité dans le cadre des nouvelles formes de menace induites par l'utilisation de l'IA
- Examen de la propagande et des opérations d'information induites par l'utilisation de l'IA
- Examen de la conduite de la guerre et des capacités infraguerrières au vu de l'utilisation de l'IA

Champ d'action 3 : aptitudes et capacités

- Intégration et utilisation renforcées de solutions basées sur l'IA par les forces armées
- Examen des possibilités de mise à niveau dans le cadre du processus d'acquisition de systèmes critiques
- Meilleure prise en compte des composants d'IA chez les fournisseurs et sous-contractants de systèmes critiques
- Examen régulier des normes technologiques des exploitants d'infrastructures critiques

Champ d'action 4 : anticipation par la collaboration, la recherche et les bancs de test

- Collaboration renforcée avec les meilleurs instituts de formation et de recherche
- Prise en compte des développements dans le domaine de l'IA dans l'image de la situation cybernétique
- Participation accrue à des organes internationaux et à des initiatives de recherche dans le domaine de l'IA
- Examen de la nécessité, du rôle et du potentiel d'un banc de test IA pour la Suisse

IA, médias et relations publiques (OFCOM. ChF, DFAE)

Champ d'action 1 : gouvernance suisse dans le domaine des intermédiaires

- Élaboration d'un rapport de gouvernance dans le domaine des intermédiaires

Champ d'action 2 : observation de l'évolution dans le domaine des médias

- Suivi de l'évolution de l'utilisation de l'IA dans le domaine des médias

Mobilité automatisée et IA (OFROU, OFAC, OFT, swisstopo, DDPS (OFPP), DETEC)

Champ d'action 1 : utilisation de l'IA dans des véhicules automatisés

- Coordination des travaux en cours concernant les véhicules automatisés
- Élaboration d'un concept de gestion du trafic aérien pour les *Unmanned Aircraft Systems* (aéronefs sans humain à bord)
- Éclaircissements concernant un projet pilote de conduite automatique en trafic ferroviaire et routier

Champ d'action 2 : échange obligatoire de données pour l'IA dans le domaine de la mobilité automatisée

- Mise en œuvre des plans de mesures existants
- Constitution d'un « réseau suisse de transport » pour la géolocalisation
- Clarifications relatives au projet pilote de communication mobile de sécurité à large bande (CMS)

Champ d'action 3 : protection des données dans le domaine de la mobilité automatisée

- Réseau de coordination avec le préposé fédéral à la protection des données et à la transparence (PFPDT)

Champ d'action 4 : réglementation et acceptation sociale de l'IA dans le domaine de la mobilité automatisée

- Homologation des véhicules automatisés
- Coordination de la procédure d'homologation des véhicules automatisés
- Clarifications relatives à la tolérance de l'IA aux pannes
- Réseau de coordination dans le domaine du droit et des affaires internationales

L'IA dans le secteur de la santé (OFSP, Swissmedic)

Champ d'action 1 : recherche sur l'être humain
- Examen des bases légales régissant le domaine du développement d'échantillons, de données et de biobanques
- Position unique sur les problématiques de l'utilisation des données et du big data

Champ d'action 2 : loi sur les produits thérapeutiques (LPTh)
- Examen de solutions possibles en matière de développement des médicaments

L'IA dans le secteur de la finance (DFF)

Champ d'action 1 : obligations en matière de comportement
- Suivi de l'évolution des obligations en matière de comportement

Champ d'action 2 : risques opérationnels
- Suivi de l'évolution des risques opérationnels au sein des établissements actifs sur les marchés financiers

Champ d'action 3 : Fixation des primes d'assurance des compagnies d'assurance privées
- Suivi des développements concernant les primes d'assurance des compagnies d'assurance privées

L'IA dans l'agriculture (OFAG)

Champ d'action 1 : conséquences de l'IA sur l'agriculture
- Suivi de l'évolution de l'IA dans l'agriculture

Énergie, climat, environnement et IA (OFEN, OFEV)

Champ d'action 1 : l'IA dans le secteur de l'énergie
- Suivi des évolutions dans le domaine « énergie »

Champ d'action 2 : l'IA dans le domaine de l'environnement et du climat
- Suivi des évolutions dans le domaine « environnement et climat »

L'IA dans l'administration (UPIC, OFS, offices fédéraux traitant de grandes quantités de données (AFD, OFS, AFC, OFAG, etc.))

Champ d'action 1 : bases de données communes
- Création et mise à disposition d'ensembles de données au sein de l'administration fédérale

Champ d'action 2 : réseau de compétences IA dans l'administration fédérale
- Clarifications approfondies en vue de la création d'un réseau de compétences IA au sein de l'administration fédérale

Champ d'action 3 : présentation des opportunités de l'IA (communication)
- Communication renforcée sur les thèmes touchant à l'IA au sein de l'administration fédérale

Champ d'action 4 : bases légales, souveraineté des données et protection des données
- Examen des bases légales relatives à l'utilisation de l'IA dans l'administration fédérale

Développement du cadre juridique général au regard de l'intelligence artificielle (DFAE)

Champ d'action 1 : mise sur pied d'un droit international spécifique à l'IA
- Élaboration d'un rapport sur l'évolution du droit international dans le domaine de l'IA

Champ d'action 2 : identification des systèmes d'IA dans l'interaction avec les consommateurs
- Suivi des développements de l'interaction des systèmes d'IA

Utilisation de l'IA dans le domaine de la justice (OFJ, DFAE)
Champ d'action 1 : utilisation de l'IA pour faciliter la prise de décision dans les domaines de l'administration, de la justice et du conseil juridique (analyse prédictive) • Observation des évolutions dans le cadre de l'aide apportée par l'IA à la prise de décision dans les domaines de l'administration, de la justice et du conseil juridique (analyse prédictive)
IA, données et droits de propriété intellectuelle (OFCOM, OFJ, IPI, SG-DFI/OFS)
Champ d'action 1 : politique relative aux données • Poursuite des travaux en cours sur la politique de la Confédération en matière de données **Champ d'action 2 : protection des données** • Poursuite des travaux en cours sur la protection des données **Champ d'action 3 : propriété intellectuelle** • Poursuite des travaux en cours sur la propriété intellectuelle

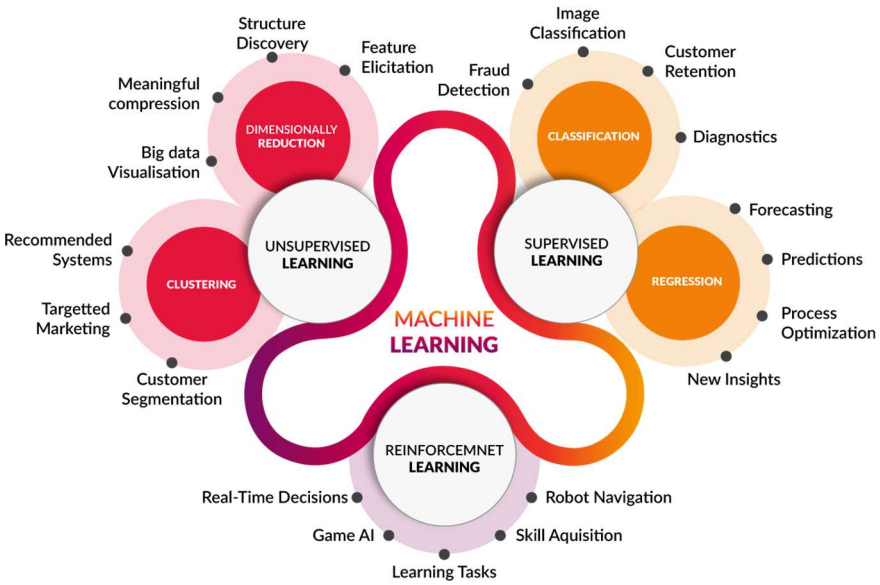
Annexe 2 : Apprentissage automatique

Cette annexe présente non seulement l'apprentissage automatique, mais aussi les notions centrales suivantes : l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage par renforcement, l'apprentissage profond, les réseaux de neurones artificiels, les réseaux de neurones profonds, les réseaux de neurones convolutifs, les réseaux de neurones récurrents et les réseaux antagonistes génératifs.

La notion d'**apprentissage automatique** désigne les approches dans le cadre desquelles des machines ont la capacité d'élargir leur propre savoir en extrayant des modèles à partir de données brutes et en établissant sur cette base des prévisions automatisées solides sous forme de données complexes.[123]

Les méthodes d'apprentissage
On distingue actuellement trois méthodes d'apprentissage principales par ces systèmes : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement. Ces approches présentent toutes des avantages et des inconvénients et se prêtent donc à différentes utilisations (Figure 14).

Figure 14 : Méthodes d'apprentissage, fonctions et domaines d'utilisation de l'apprentissage automatique



Source : "Towards Data Science: Coding Deep Learning For Beginners", disponible à l'adresse https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d

Apprentissage supervisé *(supervised learning)* : Le but de l'apprentissage supervisé est d'imiter un comportement donné à partir d'exemples. En se basant sur un jeu de données d'entraînement de paires entrée/sortie connues (p. ex. des images accompagnées d'annotations indiquant les objets

[123] Les définitions et descriptions figurant dans la présente annexe s'appuient sur les travaux suivants:
- Matt Taddy (2018), *The Technological Elements of Artificial Intelligence*, chapitre de l'ouvrage du NBER *The Economics of Artificial Intelligence: An Agenda*, Ajay K. Agrawal, Joshua Gans, et Avi Goldfarb, editors;
- Ian Goodfellow, Yoshua Bengio et Aaron Courville (2016), *Deep Learning*, MIT Press, http://www.deeplearningbook.org
- IEC White Paper: *Artificial intelligence across industries*; International Electrotechnical Commission, 2018, https://basecamp.iec.ch/download/iec-white-paper-artificial-intelligence-across-industries-en/
- SATW Technology Outlook 2019
- Yann LeCun, Yoshua Bengio & Geoffrey Hinton (2015), *Deep Learning*, *Nature* vol. 521:436-44, mai 2015.

qu'on y voit), on identifie des dépendances permettant de prédire de nouvelles données (non annotées) que l'on reconnaît donc p. ex. sur de nouvelles photos).

Applications typiques : classification d'images, filtres de spams, diagnostics médicaux

Apprentissage non supervisé *(unsupervised learning)* : Pas d'objectif didactique explicite dans les données ; l'approche permet de chercher soi-même des modèles (p. ex. des groupes) parmi des données. Les algorithmes déterminent la structure fondamentale du jeu de données, sans informations sur les critères recherchés. Un tel algorithme pourrait par exemple détecter dans une série d'images que les objets qui y sont représentés ne sont pas identiques. Il sera ensuite possible de constituer différentes catégories sans connaître les objets.

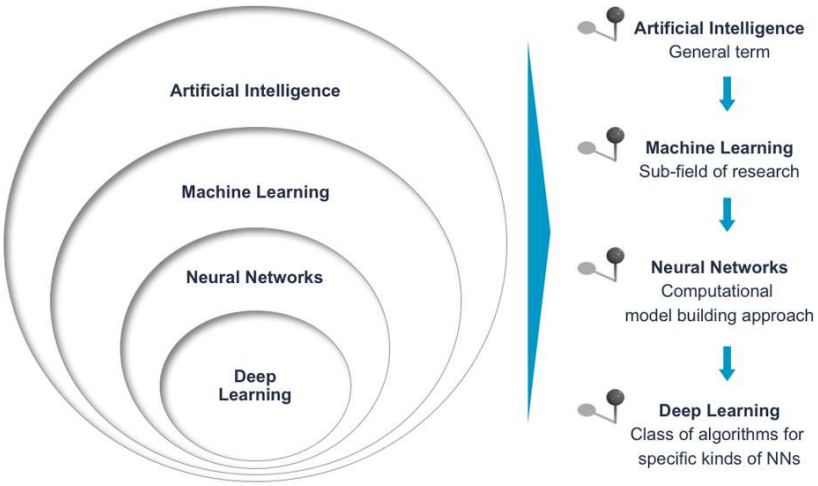
Applications typiques : segmentation de la clientèle ou recommandations concernant des produits

Apprentissage par renforcement *(reinforcement learning)* : Avec cette méthode, le système n'apprend pas à partir de données, mais via l'interaction avec son environnement (typiquement dans le cadre de simulations informatiques). Il teste différentes solutions de façon autonome et reçoit de son environnement des *retours* (feed-back), qui récompensent ou punissent les différentes façons de procéder. Il ne sait toutefois pas quelle est la meilleure façon d'agir dans telle ou telle situation. C'est davantage la répétition des actions et des retours qui permet d'apprendre les conséquences d'actions dans certaines situations.

Applications typiques : IA dans les jeux, pilotage robotisé

Apprentissage profond *(deep learning)*
Dans le cadre de l'apprentissage automatique, il existe de nombreux modèles et algorithmes qui se prêtent à différentes utilisations. Les applications actuellement les plus performantes dans le domaine de l'IA sont toutefois surtout à mettre au crédit d'une certaine catégorie de méthodes d'apprentissage automatique appelée **apprentissage profond**, qui s'appuie sur des **réseaux de neurones artificiels** (Figure 15).

Figure 15 : Les différents niveaux d'abstraction de l'intelligence artificielle

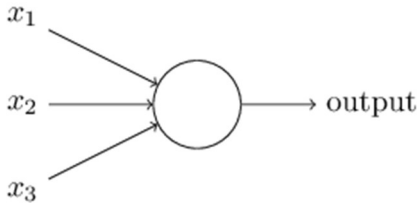


Source : Capgemini: Artificial Intelligence, Machine Learning und Data Science: Same same but different?!, disponible à l'adresse https://www.capgemini.com/de-de/2017/09/artifical-intelligence-machine-learning-und-data-science-same-same-but-different/
Voir également : Ian Goodfellow, Yoshua Bengio et Aaron Courville – 2016, *Deep Learning*, MIT Press

Les **réseaux de neurones artificiels** constituent le fondement des algorithmes actuellement utilisés dans le domaine de l'apprentissage profond. Ces réseaux sont modélisés en s'inspirant vaguement des neurones du cerveau. Ils sont typiquement composés de couches de nœuds (neurones artificiels) reliés entre elles par des poids de connexion variables (synapses).

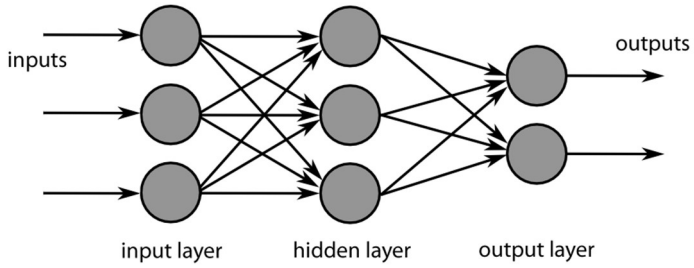
Les nœuds sont les éléments centraux de ces réseaux. Dans leur forme la plus simple, ils réceptionnent des informations et, après avoir pondéré celles-ci, décident s'il en résulte un événement (cf. Figure 16). Le choix de se rendre au travail à vélo peut p. ex. dépendre de deux facteurs : (a) du jour de la semaine (jour de travail ou pas) et (b) de la météo (beau ou mauvais temps). Avec le temps, le système *apprend* que l'importance du jour de la semaine est largement supérieure à celle de la météo et adapte la pondération en conséquence.

Figure 16 : Représentation simplifiée d'un nœud d'un réseau de neurones artificiels



Un KNN comprend de très nombreux nœuds (souvent des millions) superposés par couches. Chaque nœud d'une couche donnée est relié à plusieurs nœuds (ou tous les nœuds) de la couche suivante et peut ainsi transmettre ses informations (Figure 17). L'information, par exemple le fait d'être allé au travail à vélo, peut donc être importante pour de nombreuses autres décisions, comme celle de dîner dans un parc ou au restaurant.

Figure 17 : Représentation simplifiée d'un réseau de neurones artificiels



La qualité des applications en matière d'apprentissage automatique dépend de la possibilité de bien définir l'output à l'aide des caractéristiques (features) existantes (représentation). Le choix des bonnes caractéristiques est une tâche longue et complexe qui nécessite un savoir-faire hautement spécialisé. Souvent, cela s'avère quasiment impossible. Il est certes possible de décrire facilement une voiture à l'aide de caractéristiques telles que les pneus, les vitres, les rétroviseurs, etc., mais pratiquement impossible pour un être humain de décrire des concepts aussi abstraits en se fiant aux pixels d'une photo (dont l'ordinateur a besoin).

Un sous-domaine de l'apprentissage automatique appelé **apprentissage de représentation** résout ce problème, dans la mesure où l'apprentissage automatique porte non seulement sur le lien entre les caractéristiques et le résultat (p. ex. entre la présence de pneus et une voiture), mais aussi sur les caractéristiques nécessaires pour opérer une classification.

L'apprentissage profond franchit une étape supplémentaire en introduisant et en apprenant non seulement les caractéristiques, mais également sa propre *hiérarchie* de représentations: les représentations complexes sont créées et définies à l'aide de représentations toujours plus simples. Un tel système peut donc élaborer ses propres représentations et déterminer lui-même quelles sont les caractéristiques nécessaires et pertinentes pour la représentation en fonction de l'étape et pour l'output (Figure 18).

Figure 18 : Apprentissage automatique/profond vs systèmes basés sur des règles

		Décris-moi ce qui figure sur la photo
	Montre-moi tous les visages sur la photo	Output
Montre-moi toutes les photos dont les couleurs sont similaires	Output	Comparaison avec les caractéristiques «apprises»
Output	Comparaison avec les caractéristiques «apprises»	Transformation en caractéristiques complexes
Un «programme bien ficelé»	Algorithme d'extraction des caractéristiques imposé pour des caractéristiques prédéfinies	Extraire des caractéristiques simples
Input	Input	Input
Systèmes basés sur des règles	Apprentissage automatique	Apprentissage profond

Remarque : les encadrés grisés mettent en évidence les éléments que l'on peut apprendre à partir de données.
Source : Jaxenter: Maschinelle Bilderkennung mit Big Data und Deep Learning, disponible à l'adresse https://jaxenter.de/big-data-bildanalyse-50313, d'après Ian Goodfellow, Yoshua Bengio et Aaron Courville (2016), Deep Learning, MIT Press.

Les réseaux de neurones profonds sont les modèles centraux en matière d'apprentissage profond. Ils représentent en fin de compte la même chose que les réseaux de neurones artificiels, mais contrairement à ceux-ci, ils présentent plusieurs niveaux cachés entre les couches d'entrée et de sortie (c'est-à-dire plus qu'un « hidden layer », voir Figure 17). De tels systèmes peuvent comporter de nos jours des milliards de neurones distribués sur des douzaines de couches.

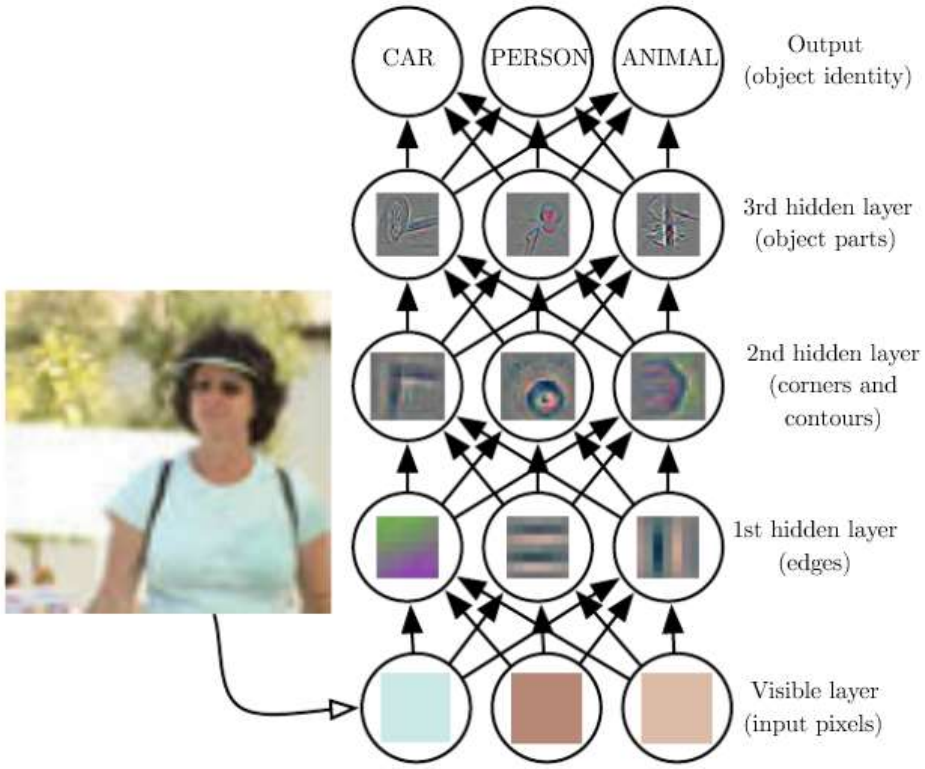
Ces couches sont « cachées » car leurs valeurs ne sont pas indiquées dans les données et ne doivent pas être imposées.[124] Au lieu de cela, le système détermine par lui-même quels concepts peuvent se révéler utiles pour expliquer les liens au sein des données observées.

Figure 19 montre comment un système d'apprentissage profond peut saisir le concept d'une photo de personne en subdivisant une classification compliquée en une série de rapports simples, emboîtés les uns dans les autres et présentant une hiérarchie propre.

L'entrée d'une photo au niveau des pixels est d'abord appréhendée à l'aide de concepts très simples. Dans un premier temps, seuls les bords sont identifiés, en comparant avec la clarté des pixels voisins. En fonction de la description des bords par le premier niveau, le second niveau cherche ensuite des concepts plus abstraits, par exemple des coins et des contours. Le troisième niveau peut à ce stade reconnaître déjà des parties entières d'objets, en réunissant différents coins et contours. Au final, cette information peut servir à identifier les parties d'objets figurant sur la photo. À la fin du processus, un tel système classe les objets représentés sur la photo.

[124] Google utilise par exemple un réseau de neurones artificiel à 30 couches pour l'analyse d'images dans *Google Photo*. https://www.androidcentral.com/google-building-deep-neural-networks-help-improve-their-search-engine

Figure 19 : Illustration d'un modèle d'apprentissage profond



Source : Ian Goodfellow, Yoshua Bengio et Aaron Courville (2016), Deep Learning, MIT Press.

Avec chaque exemple d'entraînement, le système reçoit un feedback sur sa performance de reconnaissance. Les innombrables paramètres sont ensuite adaptés. De cette manière, le système apprend peu à peu ce qui est essentiel dans le visage humain. L'entraînement se répète autant de fois qu'il faut pour que le taux d'erreur atteigne un niveau acceptable. Après cette phase d'entraînement, le système devrait être en mesure de procéder à la reconnaissance dans des données nouvelles et inconnues, avec les paramètres existants (qui ne sont dès lors plus modifiés). L'opération requiert alors une capacité de calcul bien moindre.

Parmi les innombrables variantes de réseaux de neurones profonds, nous nous contenterons d'en mettre en évidence trois:

Réseaux de neurones convolutifs : Ces réseaux sont spécifiquement conçus pour gérer des données (majoritairement) bidimensionnelles; ils se prêtent donc très bien au traitement de photos. Cette variante très performante des réseaux de neurones artificiels est aujourd'hui largement majoritaire dans les domaines de l'analyse d'images, de l'identification de personnes, de la robotique autonome, mais aussi de la reconnaissance vocale, où elle permet d'obtenir des résultats spectaculaires. Les architectures de réseaux de neurones convolutifs les plus récentes disposent de 10 à 20 couches, de centaines de millions de poids de connexion et de milliards de connexions entre les nœuds. Il y a deux ans à peine, l'entraînement de réseaux aussi gros pouvait durer des semaines, mais les progrès réalisés au niveau de la parallélisation entre matériel informatique, logiciels et algorithmes ont permis de réduire la durée de l'entraînement à quelques heures.

Réseaux de neurones récurrents : Les réseaux de neurones récurrents constituent une autre évolution importante. Tandis que les réseaux de neurones convolutifs se prêtent très bien au traitement de données bidimensionnelles, les réseaux de neurones récurrents sont spécialement conçus pour gérer les données séquentielles. Ces modèles donnent donc d'excellents résultats dans le domaine du traitement de texte. Ce que l'on appelle la «long short-term memory» (longue mémoire à court terme), développée à **l'Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA) de Lugano**, constitue l'une des implémentations les plus réussies de réseaux de neurones récurrents. Les algorithmes

basés sur ce système sont aujourd’hui utilisés quotidiennement par quelque 3 milliards de smartphones, notamment dans les domaines de la reconnaissance vocale et des traductions sur Google, et des traductions sur Facebook.

Réseaux antagonistes génératifs : Enfin, le développement des réseaux antagonistes génératifs constitue un autre jalon important dans l’évolution de l’apprentissage automatique. L’idée sous-jacente à ces réseaux est relativement simple: il s’agit de mettre en concurrence deux réseaux de neurones artificiels: l’un essaie, sur la base d’un jeu de données défini, de générer de nouvelles données qui ne se distinguent en rien des données fournies, tandis que l’autre évalue les données et tente de distinguer celles qui sont nouvelles. Cette technique apprend à générer de nouvelles données à l’aide des mêmes statistiques que celles utilisées par le jeu de données. Un réseau antagoniste génératif bien entraîné peut par exemple générer de nouvelles photos qui, pour un observateur humain, semblent parfaitement authentiques et présentent de nombreuses propriétés réalistes. Les champs d’application des réseaux antagonistes génératifs sont assez étendus; ces derniers se sont notamment fait connaître par le grand public à la suite de photos et vidéos truquées («deep fakes»).

Annexe 3 : Bibliographie

Chapitres 1 à 5

Ackerman E. *Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms*, IEEE Spectrum, 2017. https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms

Agrawal Ajay, Gans Joshua et Goldfarb Avi. « Prediction, Judgment, and Complexity: A Theory of Decision Making and Artificial Intelligence », in *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Inc., 2018.

Gopala K Anumanchipalli, Josh Chartier et Edward F. Chang. « Speech synthesis from neural decoding of spoken sentences », *Nature*, volume 568, pages 493-498, 2019. https://techcrunch.com/2019/04/24/scientists-pull-speech-directly-from-the-brain/

ASGARD. « The European Artificial Intelligence Landscape », 2017. https://asgard.vc/the-european-artificial-intelligence-landscape-more-than-400-ai-companies-made-in-europe/

Berkeley Haas. "Minority homebuyers face widespread statistical lending discrimination, study finds", 2018. https://newsroom.haas.berkeley.edu/minority-homebuyers-face-widespread-statistical-lending-discrimination-study-finds/

Conseil fédéral. *Une politique industrielle pour la Suisse*, rapport rédigé le 16 avril 2014 en réponse au postulat Bischof.

Conseil fédéral. *Rapport sur les principales conditions-cadre pour l'économie numérique*, 2017. https://www.seco.admin.ch/seco/fr/home/wirtschaftslage---wirtschaftspolitik/wirschaftspolitik/digitalisierung.html

Conseil fédéral. Rapport en réponse aux postulats 15.3854 Reynard *Automatisation. Risques et opportunités* du 16 septembre 2015 et 17.3222 Derder *Économie numérique. Identifier les emplois de demain et la manière de stimuler leur émergence en Suisse* du 17 mars 2017. https://www.seco.admin.ch/seco/fr/home/wirtschaftslage---wirtschaftspolitik/wirschaftspolitik/digitalisierung.html

Conseil fédéral. *Vision d'ensemble de la politique d'innovation*, rapport rédigé en réponse au postulat Derder 13.3073 du 13 mars 2013, 2018.

Conseil fédéral. *Bases juridiques pour la distributed ledger technology et la blockchain en Suisse*, rapport, 2018.

Carlini et Wagner. *Audio Adversarial Examples: Targeted Attacks on Speech-to-Text*, 2018, arXiv:1801.01944 [cs.LG]. https://arxiv.org/abs/1801.01944

CNN Business. "IBM's fast-talking AI machine just lost to a human champion in a live debate", 2019. https://edition.cnn.com/2019/02/11/tech/ai-versus-human-ibm-debate/index.html

Iain M. Cockburn, Rebecca Henderson, Scott Stern. *The Impact of Artificial Intelligence on Innovation, An Exploratory Analysis*, (informations bibliographiques) (téléchargement), version du 10 janvier 2018 (document de travail).

Conseil de l'Europe DGI (2017)12. *Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires*, 2017, p. 29 ss.

Conseil de l'Europe. *Discrimination, intelligence artificielle et décisions algorithmiques*, 2018.

EconSight. *Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung*, 2019. https://www.econsight.ch/artificial-intelligence/

Elsevier. « AI Report », 2018. https://public.tableau.com/profile/isabella.cingolani1149#!/vizhome/ElseviersAIprogramme/Dashboard?publish=yes

Scott Fortmann-Roe. « Understanding the Bias-Variance Tradeoff », 2012. http://scott.fortmann-roe.com/docs/BiasVariance.html

Ian Goodfellow, Yoshua Bengio et Aaron Courville. «*Deep Learning*, MIT Press, 2016. http://www.deeplearningbook.org

The Guardian. "Women must act now, or male-designed robots will take over our lives", 2018. https://www.theguardian.com/commentisfree/2018/mar/13/women-robots-ai-male-artificial-intelligence-automation

Jovanovic, Rousseau. « General Purpose Technologies, Handbook of Economic Growth », in Aghion, Durlauf (éd.), *Handbook of Economic Growth*, Elsevier, 2005, pp. 1181-1224.

Kaplan Andreas, Michael Haenlein. « Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence », *Business Horizons*, 62(1), 2018.

Regina Kiener, Walter Kälin et Judith Wyttenbach. *Grundrechte*, 3e édition, 2018.

Sebastian Lapuschkin, Stephan Wäldchen, Alexander Binder, Grégoire Montavon, Wojciech Samek et Klaus-Robert Müller. « Unmasking Clever Hans predictors and assessing what machines really learn », *Nature Communications*, volume 10, article number : 1096 (2019), 2019.

Yann LeCun, Yoshua Bengio et Geoffrey Hinton. « Deep Learning », *Nature*, volume 521:436-44, mai 2015.

J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon. « A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence », 1955. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

MIT Technology Review. "The Dark Secret at the Heart of AI", 2017. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

Sendhil Mullainathan et Jann Spiess. « Machine Learning: An Applied Econometric Approach », *Journal of Economic Perspectives*, volume 31, n° 2, printemps 2017, pp. 87-106.

OCDE. Science, Technology and Industry Scoreboard 2017 – The Digital Transformation, 2017. https://www.oecd.org/sti/oecd-science-technology-and-industry-scoreboard-20725345.htm

OCDE. *Artificial Intelligence in Society*, 2019. https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm

Roland Berger. *Artificial Intelligence, A strategy for European startups*, 2018. https://www.rolandberger.com/publications/publication_pdf/roland_berger_ai_strategy_for_european_startups.pdf

SATW *Technology Outlook 2019.* https://www.satw.ch/fr/identification-precoce/technologies/

SEFRI. *Défis de la numérisation pour la formation et la recherche en Suisse*, 2017. https://www.sbfi.admin.ch/dam/sbfi/fr/dokumente/webshop/2017/bericht-digitalisierung.pdf.download.pdf/bericht_digitalisierung_f.pdf

Strickland. « IBM Watson, heal thyself : How IBM overpromised and underdelivered on AI health care », IEEE Spectrum, vol. 56, n° 4, pp 24-31, avril 2019.

Su, Vargas et Kouichi, *One pixel attack for fooling deep neural networks*, 2017, arXiv:1710.08864 [cs.LG]. https://arxiv.org/abs/1710.08864

Szegedy et al., *Intriguing properties of neural networks*, 2014, arXiv:1312.6199v4 [cs.CV]. https://arxiv.org/abs/1312.6199v4

Matt Taddy. « The Technological Elements of Artificial Intelligence », 2018, chapitre de : Ajay K. Agrawal, Joshua Gans et Avi Goldfarb (éd.). *The Economics of Artificial Intelligence: An Agenda*, NBER, à paraître.

TA-SWISS (éd.). *Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz*, manuscrit non publié.

Rodney Brooks. « The Seven Deadly Sins of AI Predictions », MIT Technology Review, octobre 2017.

Manuel Trajtenberg. *AI as the Next GPT: A Political-Economy Perspective*, version du 23 janvier 2018 (document de travail).

Turing. « Computing Machinery and Intelligence », *Mind 49*, pp 433-460, 1950.

VALUER. « The best AI startups in Europe », 2018. https://valuer.ai/blog/the-best-ai-startups-in-europe/

Weizenbaum. « ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine », 1966. http://www.cse.buffalo.edu/~rapaport/572/S02/weizenbaum.eliza.1966.pdf

Yuan et al. *CommanderSong: A Systematic Approach for Practical Adversarial Voice Recognition*, 2018, arXiv:1801.08535 [cs.CR]. https://arxiv.org/abs/1801.08535; https://www.ics.uci.edu/~alfchen/yulong_ccs19.pdf

Instances internationales et IA

Rapport du groupe de projet *Instances internationales et intelligence artificielle*, août 2019. www.sbfi.admin.ch/ai-f

Programme pour une Europe numérique

Commission européenne. «Künstliche Intelligenz für Europa», 2018.
https://ec.europa.eu/transparency/regdoc/rep/1/2018/DE/COM-2018-237-F1-DE-MAIN-PART-1.PDF

SEFRI. *Résultats du sondage sur le programme pour une Europe numérique*, 2019.
https://www.sbfi.admin.ch/dam/sbfi/fr/dokumente/2019/07/ergebnisse-dep.pdf.download.pdf/bericht_dep_f.pdf

Changements dans le monde du travail

Conseil fédéral. *Rapport sur les principales conditions-cadre pour l'économie numérique*, 2017.
https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirschaftspolitik/digitalisierung.html

Conseil fédéral. *Conséquences de la numérisation sur l'emploi et les conditions de travail : risques et opportunités*, 2017.
https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirschaftspolitik/digitalisierung.html

Conseil fédéral. *Rapport présentant les résultats de l'enquête « Test de compatibilité numérique » – Examen des obstacles que la réglementation pose à la numérisation*, 2018.
https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirschaftspolitik/digitalisierung.html

OCDE. *The Future of Work – Employment Outlook 2019*, 2019.
https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2019_9ee00155-en

L'IA dans l'industrie et les services

SATW. *Künstliche Intelligenz in Industrie und Dienstleistungen*, rapport rédigé sur mandat du Secrétariat d'État à la formation, à la recherche et à l'innovation, 2019.
www.sbfi.admin.ch/ai-f

L'IA dans la formation

educa.ch. *Données dans l'éducation – données pour l'éducation. Bases et pistes de réflexion d'une poli-tique d'utilisation de données pour l'espace suisse de formation*. Berne. 2019.

SEFRI. « L'intelligence artificielle dans la formation », 2019. www.sbfi.admin.ch/ai-f

Tuomi I. *The Impact of Artificial Intelligence on Learning, Teaching, and Education. Policies for the future*, éd. Cabrera M., Vuorikari R. et Punie Y., Office des publications de l'Union européenne, Luxembourg, 2018.

L'intelligence artificielle dans la science et la recherche

SATW. *Künstliche Intelligenz in Wissenschaft und Forschung*, rapport rédigé sur mandat du Secrétariat d'État à la formation, à la recherche et à l'innovation, 2019.
www.sbfi.admin.ch/ai-f

L'intelligence artificielle dans la cybersécurité et la politique de sécurité

Rapport du groupe de projet *Künstliche Intelligenz in der Cybersicherheit und Sicherheitspolitik*, août 2019. www.sbfi.admin.ch/ai-f

EPFZ / CSS. « Studie KI und Sicherheitspolitik - Künstliche Intelligenz, technologischer Wandel und nationale und internationale Sicherheitspolitik », 2019.

EPFZ / CSS. « Policy Perspectives - Ein neutraler Hub für KI-Forschung », 2019.

EPFL. « Étude Cybersécurité et Politique de sécurité », 2019.

OCDE. « Artificial Intelligence in Society », 2019 https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm

IA, médias et sphère publique

Rapport du groupe de projet *Intelligence artificielle, médias et sphère publique*, août 2019. www.sbfi.admin.ch/ai-f

Dreyer et Schulz. *Künstliche Intelligenz, Intermediäre und Öffentlichkeit*, rapport à l'OFCOM établi par l'institut Alexander von Humboldt Institut für Internet und Gesellschaft (HIIG) et le Leibnitz-Institut für Medienforschung │ Hans-Bredow-Institut (HBI), 2019.

Commission fédérale des médias. Spécificités des médias à l'ère du numérique : Options d'organisation pour un paysage suisse des médias performant d'un point de vue économique et social, Bienne, 2018. https://www.emek.admin.ch/inhalte/dokumentation/22.01.2018_Besonderheiten_von_Medien_im_digitalen_Zeitalter/F_Medias_a_l_ere_numerique_22.01.18.pdf, vérifié le 13 août 2019.

Fichter. « Die Schweiz wappnet sich für den Angriff aus dem Silicon Valley », Republik, 16 mai 2018. https://www.republik.ch/2018/05/16/die-schweiz-wappnet-sichfuer-den-angriff-aus-dem-silicon-valley, vérifié le 13 août 2019.

Gillespie, Tarleton. #trendingistrending. Wenn Algorithmen zur Kultur werden, in: Robert Seyfert et Jonathan Roberge (éd.), Algorithmuskulturen. Über die rechnerische Konstruktion der Wirklichkeit, Bielefeld, éd. transcript Verlag, pp 75–106, 2017.

Goldhammer, Dietrich, Prien. Künstliche Intelligenz, Medien und Öffentlichkeit. Wissenschaftlicher Bericht, Berlin, 2019.

Jarren. Normbildende Macht, in: epd medien (24), pp 35–39, 2018b.

Livingstone. Audiences in the Age of Datafication : Critical Questions for Media Research, in: Television & New Media 20 (2), pp 170–183, 2019.

Lobigs, Neuberger. Meinungsmacht im Internet und die Digitalstrategien von Medienunternehmen. Neue Machtverhältnisse trotz expandierender Internet-Geschäfte der traditionellen Massenmedien-Konzerne. Gutachten für die Kommission zur Ermittlung der Konzentration im Medienbereich (KEK), Leipzig, 2018.

Saurwein, Just, Latzer. Algorithmische Selektion im Internet: Risiken und Governance automatisierter Auswahlprozesse, in: kommunikation@gesellschaft, 22 pages, 2017.

Conseil fédéral suisse. Un cadre juridique pour les médias sociaux : nouvel état des lieux. Rapport complémentaire du Conseil fédéral sur le postulat Amherd 11.3912 « Cadre juridique pour les médias sociaux », Berne, 2017. https://www.bakom.admin.ch/dam/bakom/fr/dokumente/informationsgesellschaft/social_media/social%20media%20bericht.pdf.download.pdf/rapport-media-sociaux-2017-FR.pdf, vérifié le 13 août 2019.

Conseil suisse de la presse (éd.). Code déontologique. https://presserat.ch/fr/code-de-deontologie-des-journalistes/erklaerungen/, vérifié le 13 août 2019

Mobilité automatisée et IA

Rapport du groupe de projet *Mobilité automatisée et intelligence artificielle*, août 2019. www.sbfi.admin.ch/ai-f

Conseil fédéral. *Conduite automatisée – Conséquences et effets sur la politique des transports*, rapport du Conseil fédéral en réponse au postulat Leutenegger Oberholzer 14.4169 « Automobilité. Voitures sans conducteur. Impact pour la politique des transports », 2016.

L'intelligence artificielle dans la finance

Su, Vargas et Kouichi, *One pixel attack for fooling deep neural networks*, 2017, arXiv:1710.08864 [cs.LG]. https://arxiv.org/abs/1710.08864

L'intelligence artificielle dans l'énergie, le climat et l'environnement

Rapport du Conseil fédéral. « Environnement Suisse 2018 », 2018. https://www.bafu.admin.ch/bafu/fr/home/etat/publications-etat-de-l-environnement/environnement-suisse-2018.html

Office fédéral de l'énergie. « La digitalisation du monde de l'énergie - Dialogpapier zum Transformationsprozess », 2019. https://www.bfe.admin.ch/bfe/fr/home/approvisionnement/digitalisation.html

Office fédéral de l'énergie. Programme pilote, de démonstration et programme phare, 2019. https://www.bfe.admin.ch/bfe/fr/home/recherche-et-cleantech/programme-pilote-de-demonstration-et-programme-phare.html

L'intelligence artificielle dans l'administration

Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence. «Improving refugee integration through data-driven algorithmic assignment», *Science*, Vol. 359, Issue 6373, 2018, pp 325-329. https://science.sciencemag.org/content/359/6373/325

Résultats de l'enquête sur l'utilisation de l'intelligence artificielle (IA) par les cantons du 15 juillet 2019, canton de Lucerne, manuscrit non publié.

EPF Zurich. « Algorithmus verbessert Erwerbschancen von Flüchtlingen », 2018. https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2018/01/algorithmus-verbessert-erwerbschancen-von-fluechtlingen.htm

L'intelligence artificielle dans la justice

Commission européenne pour l'efficacité de la justice. « Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement », 2018 https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b