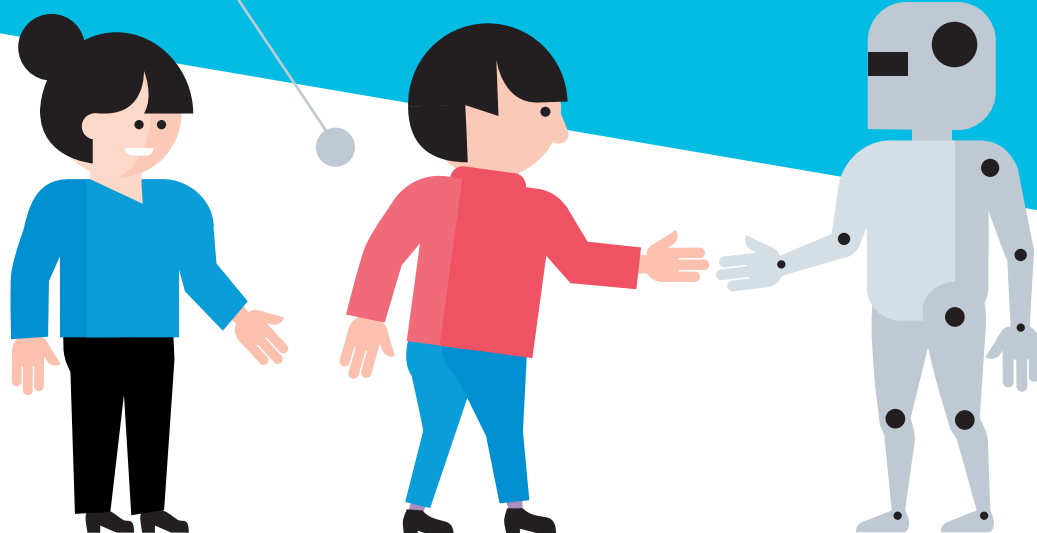


Sans neurones mais intelligent

Synthèse de l'étude «Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz» de TA-SWISS



TA-SWISS, Fondation pour l'évaluation des choix technologiques et centre de compétence des Académies suisses des sciences, entend mener une réflexion sur les répercussions – opportunités et risques – de l'utilisation de nouvelles technologies.

Cette synthèse se base sur une étude scientifique réalisée pour le compte de TASWISS par un groupe de projet interdisciplinaire dirigé par Markus Christen (Digital Society Initiative de l'Université de Zurich), Clemens Mader (Laboratoire fédéral d'essai des matériaux et de recherche Empa) et Johann Čas (Académie autrichienne des sciences ÖAW). Elle présente les principaux résultats et les recommandations de l'étude sous forme condensée et s'adresse à un large public.

Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz

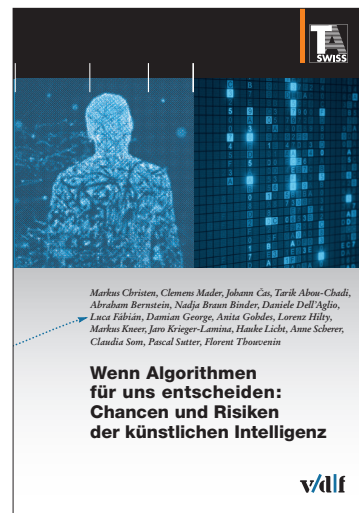
Markus Christen, Clemens Mader, Johann Čas, Tarik Abou-Chadi, Abraham Bernstein, Nadja Braun Binder, Daniele Dell'Aglio, Luca Fábíán, Damian George, Anita Gohdes, Lorenz Hilty, Markus Kneer, Jaro Krieger-Lamina, Hauke Licht, Anne Scherer, Claudia Som, Pascal Sutter, Florent Thouvenin

TA-SWISS, Stiftung für Technologiefolgen-Abschätzung (éd.)

vdf Hochschulverlag an der ETH Zürich, 2020.
ISBN 978-3-7281-4001-2

L'étude est également disponible en open access :
www.vdf.ch

La synthèse peut être téléchargée gratuitement :
www.ta-swiss.ch



L'intelligence artificielle en bref	4
Leurs chances ...	4
... leurs risques ...	4
... et quelques recommandations	4
L'intelligence artificielle, enfant de la numérisation	5
Du problème mathématique au modèle complexe	5
Surpasser l'humain	6
Distorsions, manque d'équité et manque de confiance	7
Des domaines d'application variés	8
L'IA et la consommation : une virée de shopping personnalisée	8
Les préférences personnelles servent de repères	8
Être humain ou machine ?	8
Glu numérique	9
Protéger les données c'est protéger la personne	9
Difficile équilibre entre personnalisation et protection des données	10
L'IA et la communication publique : formation de l'opinion dans la chambre d'écho des médias	10
Encore plus fort dans la bulle	10
L'usine à trolls et ses théories de conspiration	11
La démocratie en danger	11
Un regard sceptique sur les réseaux sociaux	11
L'IA dans le monde professionnel : quand les machines font le travail d'un cerveau	12
Une carrière sous le signe de l'IA	12
Travailler comme une machine	13
Des mesures contre la division du marché du travail sont requises	13
L'IA dans la formation et la recherche : l'intelligence humaine assistée	14
À chaque enfant son super professeur	14
À la recherche des talents de demain	15
Influence croissante des entreprises privées sur la formation	15
L'IA fait l'objet de recherches – et fait avancer la recherche	15
Développer les compétences clés et assurer la transparence	15
L'IA dans l'administration : des compétences souveraines pour l'intelligence artificielle ?	16
Lutte contre la fraude et la criminalité	16
Les droits fondamentaux l'emportent sur l'arbitraire de l'État	16
La Suisse reste prudente	17
Pas de délégation du pouvoir décisionnel à la machine	18
Les machines « intelligentes » – un défi pour la législation en vigueur	18
Tenir les systèmes d'IA en laisse ?	18
Créativité de première et de seconde main	18
L'équité est difficile à calculer	19
Lignes directrices en matière d'éthique au niveau européen	19
Distorsions et manque de transparence	19
Recommandations : ouvrir le champ	20
Examiner les applications spécifiques	20
Des exigences élevées à l'égard des autorités	20
Transparence et information assimilable	20
Un label de qualité au lieu d'une autorisation générale de mise sur le marché	20
Voir au-delà de la technologie	20

L'intelligence artificielle en bref

Contrairement aux programmes informatiques classiques, qui exécutent des tâches clairement spécifiées selon des étapes préétablies, le fonctionnement de l'intelligence artificielle (IA) est de nature créative. De nombreux systèmes d'IA sont capables, grâce à un entraînement sur des ensembles de données volumineux, de détecter des schémas cachés et de proposer des solutions auxquelles les utilisatrices et utilisateurs n'auraient pas pensé par eux-mêmes. Les systèmes d'IA élaborés combinent diverses compétences comme le traitement des langues, l'organisation de l'information et l'apprentissage machine, et sont même capables d'exécuter des tâches exigeantes actuellement encore confiées à des professionnels bien formés.

Leurs chances ...

Les systèmes d'IA peuvent être utilisés dans les domaines d'application les plus variés. Ils sont à bien des égards plus rapides et souvent plus précis que les êtres humains et pourraient leur permettre d'exécuter certaines tâches complexes beaucoup plus efficacement qu'auparavant.

Les systèmes d'IA permettent souvent de mieux adapter les prestations et les offres aux besoins et aux capacités de chaque individu. Cette personnalisation est un des grands atouts de l'IA.

... leurs risques ...

De nombreux systèmes d'IA doivent être entraînés avec d'énormes ensembles de données afin d'acquies les compétences souhaitées. La sphère privée et la protection des données risquent d'être fragilisées par les immenses besoins en données de l'IA.

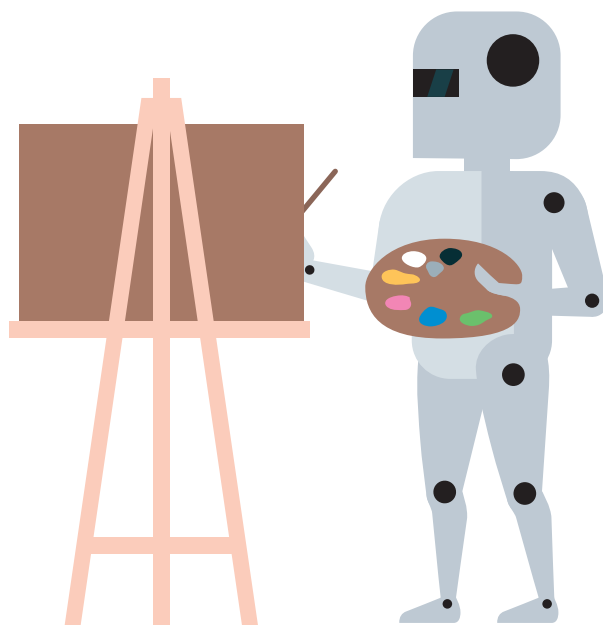
Si les ensembles de données sur lesquelles l'IA se base contiennent des erreurs, les résultats seront également erronés. Les données non équilibrées sont elles aussi problématiques, car si elles conduisent à des résultats mathématiquement corrects mais faussés en termes de contenu, certains groupes de personnes risquent d'être systématiquement désavantagés.

Les systèmes auto-apprenants de l'IA continuent de se développer et peuvent aboutir à des résultats devenus incompréhensibles pour leurs auteurs, et surtout pour les personnes concernées.

... et quelques recommandations

L'utilisation de systèmes d'IA devrait être transparente et facile à identifier afin que les gens sachent qu'ils interagissent avec un système d'IA et non avec un être humain.

Les décisions importantes qui touchent les individus ne doivent pas être déléguées à un système d'IA



sans en examiner soigneusement les avantages et les inconvénients. S'il s'agit de questions revêtant une importance pour la sphère privée, le résultat fourni par le système doit en principe être vérifié et justifié par un être humain.

Le recours à l'IA a des conséquences qui vont bien au-delà des implications techniques. Quiconque développe l'IA, l'utilise, ou encore exploite ses résultats, devrait donc également avoir des connaissances en matière d'éthique et de droit et, en outre, être à la fois disposé et capable de s'engager dans

une coopération interdisciplinaire avec des personnes représentant d'autres domaines scientifiques.

Sur le plan méthodologique, l'étude s'appuie dans un premier temps sur une vaste analyse de la littérature. Celle-ci a servi de base à une deuxième phase au cours de laquelle des spécialistes ont été invités à plusieurs reprises à évaluer les thèses et problématiques qui avaient été identifiées. Les résultats de l'analyse de la littérature et des enquêtes ont permis de formuler des recommandations à l'intention de différents destinataires.

L'intelligence artificielle, enfant de la numérisation

L'intelligence artificielle (IA) a depuis longtemps cessé d'être de la science-fiction pour faire progressivement partie de notre vie quotidienne. C'est une raison suffisante pour reconnaître non seulement ses avantages – comme l'accroissement de l'efficacité – mais aussi ses zones d'ombre, et en particulier sa soif de données.

Si l'on part de l'origine du mot « intelligence », il signifie la liberté intellectuelle des êtres vivants qui ne suivent pas de chemin prédéterminé, mais peuvent réfléchir à des alternatives et en choisir. Mais que signifie exactement IA, l'acronyme d'« intelligence artificielle » ?

La communauté scientifique a commencé à s'intéresser à l'IA en 1950. À l'époque, l'informaticien britannique Alan Turing avait déclaré dans un article du magazine *Mind* qu'on pouvait dire d'une machine qu'elle était intelligente à partir du moment où elle était capable d'imiter si parfaitement une conversation avec un être humain qu'elle donnait l'impression de dialoguer avec une personne réelle. Le test dit de Turing, effectué sur un écran avec un clavier et sans contact visuel ou auditif avec l'interlocuteur, est considéré comme un test décisif pour l'intelligence des systèmes artificiels depuis la première conférence scientifique sur l'IA tenue à Dartmouth (USA) en 1956. Le terme « intelligence artificielle » a également été adopté lors de cette conférence.

Du problème mathématique au modèle complexe

L'IA, en tant qu'objet de recherche en informatique, s'est considérablement complexifiée depuis sa conception. Tandis que l'accent était initialement mis sur les algorithmes de résolution de problèmes mathématiques et géométriques, ce n'est que dans les années 1990 que des processus plus complexes ont pu être abordés grâce à la puissance de calcul accrue des ordinateurs. Pour pouvoir extraire des modèles significatifs à partir de signaux acoustiques, d'images ou d'autres entrées, de nombreux systèmes d'IA doivent être entraînés avec une énorme quantité de données. Le développement de la numérisation et la disponibilité croissante de gros volumes de données constituent les fondements de l'IA.

Aujourd'hui, l'IA englobe divers champs de recherche scientifique dans lesquels on tente de reproduire une action rationnelle ou humaine au moyen d'un artefact, c'est-à-dire d'un programme informatique. Ainsi, un sous-groupe de travaux vise à reproduire la pensée et l'apprentissage des êtres humains. Outre les procédures qui se basent sur de grands volumes de données pour déterminer les schémas et concevoir des modèles à l'aide d'algorithmes, on utilise également des procédures qui s'appuient sur les connaissances d'experts ou les étapes de la pensée logique. Les systèmes d'IA élaborés combinent diverses compétences comme le traitement des langues, l'organisation de l'information et l'apprentissage machine.

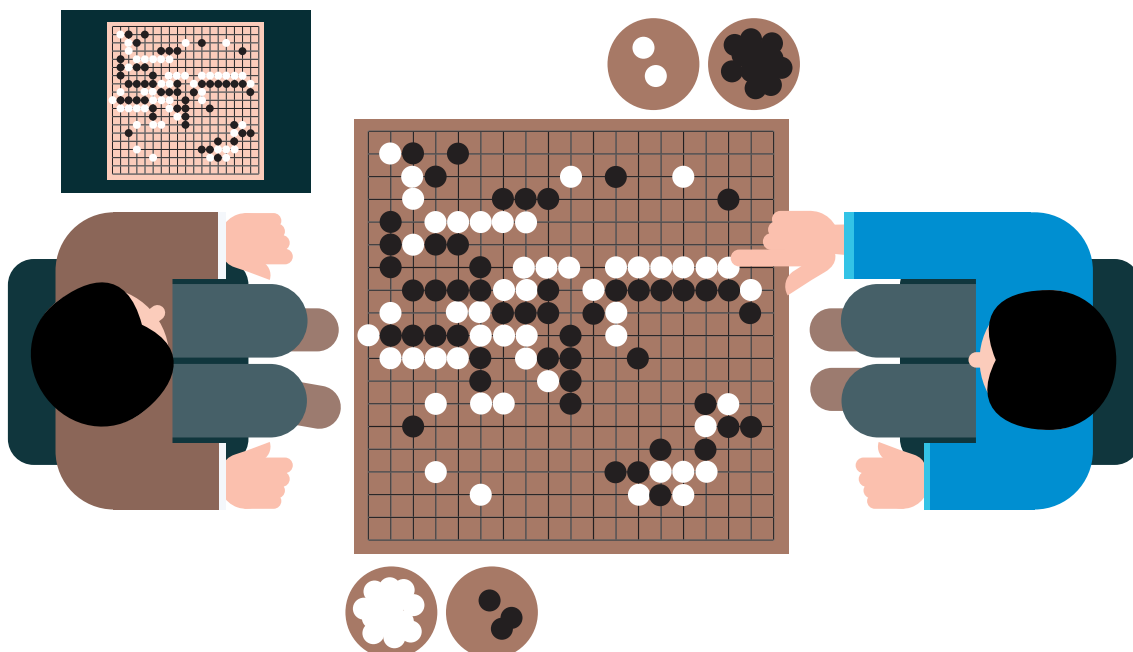
Au cours des dernières années, l'IA a beaucoup fait pour améliorer le confort et accroître la sécurité dans notre vie quotidienne : les concepts de mobilité qui combinent les transports publics avec des voitures de location ou des voitures particulières reposent sur des systèmes d'IA, les compagnies de chemin de fer augmentent leur sécurité grâce à une reconnaissance intelligente des dommages, et la production industrielle utilise l'IA pour rendre les processus plus efficaces – pour ne citer que quelques exemples. La recherche fondamentale sur la nature du cosmos, telle qu'elle est menée au CERN, serait impossible sans l'analyse à l'aide de l'IA d'énormes volumes de données ; les simulations du changement climatique font également appel à l'IA. L'utilisation de données factuelles de ce type par les systèmes d'IA ne pose en général aucun problème et se traduit souvent par des avantages considérables pour la société. Cela devient délicat lorsque l'IA exploite des données à caractère personnel. L'étude de TA-SWISS porte également sur ce type d'applications.

Surpasser l'humain

Ce sont surtout les systèmes d'IA basés sur l'apprentissage machine et les réseaux neuronaux qui résolvent des problèmes là où les programmes informatiques classiques échouent. Dans les tâches exigeantes et complexes comme la traduction, la reconnaissance et l'analyse d'images ainsi que les jeux, l'IA est aujourd'hui non seulement capable d'atteindre un niveau équivalent à celui d'un être humain, mais parfois même de le dépasser.

Une étape importante a été franchie en mars 2016, lorsqu'une IA a battu le meilleur joueur au monde de l'époque au jeu de go, un jeu de stratégie complexe d'origine asiatique. Mais cela ne s'est pas arrêté là : par la suite, la société de logiciels DeepMind, une filiale de Google, a continué de développer son programme et a élaboré une version capable d'apprendre elle-même à jouer en se basant uniquement sur les règles du jeu au lieu de s'appuyer, comme dans les versions précédentes, sur l'analyse des coups joués par d'excellents joueurs humains dans d'innombrables parties. Commençant par des parties aléatoires de niveau débutant, le logiciel Alpha Go Zero a atteint en trois jours une capacité de jeu professionnelle et a dépassé son prédécesseur qui avait battu le numéro un mondial.

Les systèmes d'IA se développent très rapidement et sont toujours plus à même d'exécuter des tâches auparavant réservées à l'être humain ; ils sont ainsi devenus une technologie fondamentale pour d'innombrables applications. Lorsque de nouvelles formes d'IA sont utilisées, le rôle de l'être humain dans la définition et la résolution des problèmes change : jusqu'à présent, c'était une personne qui identifiait les questions et les tâches, élaborait le plan d'action requises pour résoudre chaque problème (et, si nécessaire, apportait des corrections après les premiers échecs) avant de finalement mettre en œuvre les instructions dans des domaines d'application définis d'une manière relativement précise ; si l'évaluation montrait que les mesures prises pour résoudre le problème n'avaient pas abouti, cette personne apportait alors les corrections nécessaires.



Maintenant, au contraire, l'apprentissage machine se fait en grande partie sans intervention humaine : de grands ensembles de données préstructurent la problématique ; puis les actions requises pour résoudre le problème – les algorithmes – sont développées sur la base des modèles issus des données et à l'aide d'un apprentissage intensifié de la machine. La machine effectue également le retour d'information de manière autonome et, le cas échéant, adapte et affine l'algorithme sur la base d'autres données recueillies de manière autonome ; l'être humain reste dans une large mesure en dehors du processus, même si la décision potentielle d'écarter le résultat du système d'IA lui revient.

Ce qui paraît abstrait a des conséquences tangibles dans la vie quotidienne – par exemple, en matière d'octroi de prêts par les banques : dans la procédure traditionnelle, un professionnel qualifié examine une demande de financement sur la base de différents critères et en fonction de son expérience. Le logiciel quant à lui examine le portefeuille clientèle de la banque, analyse les données personnelles et la situation financière des clientes et clients et sait en outre lesquels d'entre eux remboursent leur crédit dans les délais. À partir de ces données, le logiciel crée un modèle de prédiction pour évaluer la probabilité qu'une nouvelle cliente ou nouveau client rembourse son prêt. Les systèmes auto-apprenants qui s'améliorent constamment par l'expérience posent un problème fondamental : à qui incombe la responsabilité d'une décision si la chaîne de processus qui y a conduit demeure obscure parce que plus personne ne peut la retracer ? Même si c'est un être humain qui juge les résultats du logiciel et prend la décision finale, cela devient délicat lorsqu'il comprend de moins en moins bien son outil.

Distorsions, manque d'équité et manque de confiance

Lorsque l'IA trouve des moyens de résoudre des problèmes, un certain nombre de difficultés de fond surgissent.

Si l'IA s'appuie sur de vastes ensembles de données pour fonctionner, elle s'appuie également sur les erreurs ou les défauts cachés dans ces données. Il se peut cependant qu'une certaine partialité influence le comportement de l'algorithme, voire l'induit délibérément en erreur si les données d'apprentissage correspondantes ont été manipulées. Néanmoins, le déséquilibre qui peut se produire dans les grands ensembles de données – qui sont par exemple compilés à la suite de requêtes dans les

moteurs de recherche – n'est pas nécessairement dû à une intention frauduleuse. Au contraire, un tel déséquilibre a pu apparaître au cours du temps, à force de refléter des jugements de valeur et des habitudes jamais remises en question. Dans ces cas, l'IA peut aider à détecter de telles distorsions, à condition d'être utilisée avec la diligence requise.

Un autre problème est dû à l'indépendance croissante de l'IA. En effet, l'état initial du logiciel est connu du programmeur. Toutefois, surtout avec les réseaux dits neuronaux artificiels, la force et la pondération des différentes connexions se modifient au cours des nombreux cycles d'entraînement, de sorte que le logiciel propose finalement des solutions dont le fondement logique ou physique est difficile à identifier. En pratique, de tels algorithmes d'IA opaques, qui agissent comme une véritable boîte noire (black box), sont d'une utilité limitée si l'on souhaite comprendre comment un système tire ses conclusions.

Non seulement les données peuvent être déformées, mais les algorithmes en tant que tels également. En effet, les modèles sont créés par les développeurs qui peuvent être amenés à donner la priorité à certaines valeurs et à certains intérêts par rapport à d'autres. De tels algorithmes « inévitables » ont déjà donné des résultats biaisés, par exemple en supposant que les personnes vivant dans les zones résidentielles plus pauvres ont une plus grande propension à la délinquance.

Des études empiriques montrent que de nombreuses personnes ont davantage confiance dans une décision humaine, même lorsqu'elles savent qu'il est prouvé que l'IA permet de prendre des décisions plus objectives. Il existe toutefois aussi des études qui montrent le contraire, à savoir que les gens comptent parfois trop sur les résultats des processus de décision automatisés. Il est difficile d'évaluer correctement le niveau de confiance que l'IA mérite.

Enfin, certains signes laissent à penser que l'IA renforce la suprématie de l'économie : comme de nombreuses nouvelles formes d'IA dépendent d'énormes ensembles de données, les entreprises qui ont accès à de grandes quantités de données bénéficient d'un avantage concurrentiel. Ainsi, des entreprises technologiques de premier plan aux États-Unis et en Chine utilisent les informations concernant leurs clients pour s'impliquer activement dans le développement de l'IA. Ce phénomène bien connu dans l'économie de l'Internet, où les entreprises déjà bien positionnées se renforcent rapidement et créent ainsi des oligopoles, va probablement s'intensifier.

Des domaines d'application variés

L'IA est déjà utilisée dans des domaines très différents. L'une des forces des systèmes d'IA réside dans la capacité d'adapter les services et les offres à l'individu.

L'IA est de plus en plus utilisée dans les domaines de la consommation, des médias (traditionnels et nouveaux), du monde du travail, de la formation et de la

recherche, et de l'administration. L'étude de TA-SWISS se penche sur ces cinq domaines d'application. L'ordre choisi dans cette brochure diffère de celui du rapport d'étude détaillé. Il reflète l'expérience quotidienne de la plupart des utilisatrices et utilisateurs, qui sont plus souvent susceptibles de faire des achats en ligne et de lire des nouvelles sur leur écran que de contacter les autorités. En revanche, il ne révèle rien de l'importance de chaque domaine d'application.

L'IA et la consommation : une virée de shopping personnalisée

Quiconque commande des marchandises sur Internet se retrouve inévitablement en contact avec l'IA. L'expérience d'achat personnalisée est devenue une évidence, les processus informatiques fonctionnent discrètement en arrière-plan – et ne sont presque jamais remis en question.

« Les produits suivants peuvent également vous intéresser », « Commandez-les maintenant ! » ou « Assortissez-les » : ces formules sont bien connues des personnes qui achètent sur Internet et à qui des boutiques spécialisées signalent encore d'autres produits susceptibles d'être attrayants et pratiques pour elles. Ces systèmes de recommandation utilisent les « empreintes numériques » de la clientèle : sur la base d'une analyse à l'aide de l'IA de tous les achats effectués sur la plateforme en question et des données personnelles des clientes et clients, mais aussi de leurs autres comportements tels que le temps passé sur la page ou les mouvements de souris, le système tire des conclusions sur les préférences de certains groupes de clientèle.

La publicité devient elle aussi toujours plus personnalisée. Google et certains grands fournisseurs comme Alibaba et Amazon sont actifs dans ce domaine, de même que les réseaux sociaux tels que Facebook. Ils font de la publicité ciblée pour une grande variété de biens et de services, souvent pour le compte d'autres entreprises. La technologie qui fonctionne en arrière-plan est comparable à celle des systèmes de recommandation.

Les préférences personnelles servent de repères

Les systèmes de recommandation utilisant l'IA offrent avant tout l'avantage de pouvoir s'orienter plus facilement dans l'univers de la consommation. Et au lieu d'être largement diffusée, la publicité qui apparaît sur l'écran de la consommatrice ou du consommateur en question ne porte que sur des produits adaptés à ses intérêts.

Tout ce qui aide la clientèle à s'orienter dans le labyrinthe des marchandises aide aussi les fournisseurs à diminuer leurs dépenses. Les coûts de traitement peuvent être réduits jusqu'à 20 % grâce au recours à l'IA, et la publicité qui se concentre sur le bon groupe cible subit moins de pertes de diffusion. Le service à la clientèle atteint lui aussi un nouveau niveau et devient toujours plus individualisé. En outre, les recettes peuvent être maximisées, notamment grâce à la tarification dynamique : l'IA permet aux fournisseurs de déterminer la volonté de leur cliente ou client à dépenser et de calculer automatiquement la valeur des marchandises.

Être humain ou machine ?

L'IA est de plus en plus utilisée non seulement pour l'achat de biens mais aussi de services. En mai 2018, par exemple, Google a lancé l'assistant personnel numérique Duplex, qui utilise une voix féminine d'un réalisme trompeur pour prendre rendez-vous au salon de coiffure ou réserver une table au restaurant au nom de sa ou de son propriétaire. Parce

que Duplex a effrayé beaucoup de monde, même aux États-Unis où l'on est plutôt ouvert d'esprit en matière de technologies, Google a modifié la configuration de l'assistant de sorte qu'il se présente au début d'une conversation. Dans son blog, le groupe indique d'autres applications possibles de cette technologie, par exemple pour divers services de renseignements téléphoniques commerciaux ou administratifs : au lieu d'endurer de longues files d'attente jusqu'à ce qu'un interlocuteur personnel soit disponible, on pourrait plus souvent avoir recours à des assistants numériques pour répondre aux questions les plus fréquentes.

Duplex illustre clairement l'un des problèmes principaux de l'IA : il est de plus en plus rare qu'une personne qui commande des biens et des services par Internet ou via son téléphone mobile sache si c'est une personne en chair et en os qui lui répond ou s'il s'agit d'une interaction basée sur un algorithme. En outre, il est difficile pour une cliente ou un client de savoir si et quand les systèmes d'IA sont utilisés, mais aussi comment et dans quel but. Même si les consommatrices et consommateurs sont informés que leurs conversations ou leurs données personnelles sont enregistrées, ils ne savent pas quelles déclarations les algorithmes en extraient plus tard, ni avec quelle exactitude.

Glu numérique

Un compte sur lequel sont stockées toutes les informations sur toutes les commandes déjà effectuées par une personne ainsi que ses données individuelles est vraiment pratique : le portail d'entrée personnalisé avec les photos des derniers articles achetés crée un climat familier et, comme l'adresse et le numéro de carte de crédit sont enregistrés, la saisie fastidieuse de ces données pour chaque commande ultérieure est évitée. Au fil du temps, c'est-à-dire à chaque transaction, le système apprend à mieux connaître les goûts de sa clientèle, de sorte que les recommandations sont plus précises. Le revers de la médaille : par confort, la cliente ou le client est incité à rester chez un fournisseur, même s'il existe un assortiment plus avantageux ailleurs. Cet effet est connu parmi les experts sous le nom de « capacité de retenir le client » (stickiness).

De plus, un réseau relie également les individus du simple fait de sa taille. En effet, il est utile de faire partie d'un réseau aussi vaste que possible pour atteindre de nombreuses personnes. Il est également avantageux de pouvoir exécuter différentes tâches via une plateforme unique, de sorte qu'il ne

soit pas nécessaire de se connecter séparément pour chaque transaction ou interaction. En conséquence : les données finissent par être concentrées auprès de quelques grands groupes dont le pouvoir de marché est considérable.

Protéger les données c'est protéger la personne

Les traces que nous laissons sur Internet sont révélatrices. Le réseau social Facebook lui-même a publié une étude portant sur plus de 8 000 utilisatrices et utilisateurs volontaires dont les caractéristiques personnelles ont été évaluées de manière plus précise par un système d'IA que par leurs amis sur la base des likes obtenus. La machine n'a eu besoin que de dix likes pour donner de meilleurs résultats que les collègues ; à partir de septante likes analysés, l'évaluation de la machine a dépassé celle du cercle d'amis. D'autres études ont montré que les systèmes d'IA sont capables de reconstituer avec une précision étonnante les caractéristiques individuelles d'une personne – y compris son orientation sexuelle – à partir de sa photo de profil.

L'IA est particulièrement efficace lorsqu'elle associe des données personnelles provenant de différentes sources. C'est un avantage dont des entreprises comme Facebook et Google savent tirer profit car de nombreux services populaires sont entre leurs mains : Facebook possède notamment le service photo Instagram et la messagerie WhatsApp, tandis que Google – avec de nombreuses autres entreprises – possède le portail vidéo YouTube et la société de suivi de la condition physique Fitbit. D'autres géants de la technologie comme Amazon et Alibaba ont également acheté un grand nombre de plateformes Internet et d'ensembles de données, ce qui leur permet de relier entre elles des données provenant de différentes sources et d'accroître ainsi leur avantage concurrentiel par rapport aux entreprises plus petites. Néanmoins, les spécialistes ne voient pas forcément un danger dans les ensembles de données concentrés : comme les utilisatrices et utilisateurs peuvent mettre leurs données à la disposition de plusieurs fournisseurs, l'accès à certaines informations en tant que tel ne garantit pas une avance sur la concurrence.

Laisser à la seule cliente ou au seul client la charge de protéger ses données n'est pas une solution. Il est vrai que les conditions générales (CG) doivent mentionner les informations qui sont recueillies et réunies. Toutefois, le fait qu'une personne accepte les conditions ne permet pas pour autant d'en

déduire qu'elle autorise l'accès à ses données, car il est peu probable qu'elle ait jamais conscience de toute l'étendue des informations recueillies ni de l'utilisation qui en est faite : pour lire dans leur totalité ne serait-ce que les termes et conditions de ses services les plus fréquemment utilisés, l'utilisatrice ou l'utilisateur lambda devrait y passer en moyenne plusieurs semaines par an.

Difficile équilibre entre personnalisation et protection des données

De l'avis des personnes interrogées dans le cadre de l'étude TA, il est urgent que les utilisatrices et utilisateurs prennent davantage conscience de la valeur de leurs propres données. Cependant, il est difficile pour les entreprises de concilier personnalisation et protection des données. C'est du moins l'appréciation des expertes et experts interrogés. La menace de la perte de confiance doit également être prise au

sérieux par les fournisseurs sur Internet. Des données ouvertes avec des options de contrôle pour les clientes et les clients, une transparence quant à l'utilisation de l'IA et le droit de supprimer des données sont des mesures considérées par les spécialistes interrogés comme prometteuses pour promouvoir la confiance des clients.

L'émergence d'oligopoles, dans lesquels quelques géants dominant le marché, est citée par les personnes interrogées comme un autre danger important potentiellement associé aux systèmes d'IA. En définitive, de telles structures pourraient entraîner une réduction de l'offre et limiter la capacité des clientes et clients à changer de fournisseur. Il est donc important de veiller à ce qu'ils puissent emporter leurs données personnelles avec eux lorsqu'ils changent de fournisseur. Il est également important d'observer attentivement les évolutions du marché et de disposer d'options de contrôle et d'intervention qui garantissent la libre concurrence.

L'IA et la communication publique : formation de l'opinion dans la chambre d'écho des médias

Au cours des dix dernières années, la presse, la radio et la télévision ont perdu de leur importance : de plus en plus de personnes s'informent via les médias sociaux tels que Facebook ou Twitter. Les informations basées sur des enquêtes professionnelles et placées dans un contexte plus large se retrouvent en concurrence avec les blogs ou les posts sur les réseaux sociaux.

En 2017, 83 % de la population suisse s'informait en ligne, dont 45 % via les médias sociaux. Outre Facebook et Twitter, les plateformes telles que Google News et Apple News, qui regroupent des contenus provenant d'autres sources, sont toujours plus consultées. En conséquence : un nombre croissant d'utilisatrices et d'utilisateurs partagent des articles et des informations directement les uns avec les autres et retransmettent souvent des nouvelles provenant de fournisseurs traditionnels. La source des informations relayées devient toujours plus floue.

Encore plus fort dans la bulle

Les médias traditionnels sont supposés fournir une information équilibrée sur un large éventail de sujets et ainsi refléter la diversité d'opinions. Toute personne qui s'informe via la presse ou la radio et la télévision apprend inévitablement aussi des choses qui l'intéressent moins.

La situation est différente lorsque les informations sont partagées via les réseaux sociaux. Lorsque que le compte d'un ou d'une titulaire reçoit de nombreux fils d'actualité (newsfeed), l'algorithme détermine quel contenu est affiché en premier en fonction des informations recueillies sur cette personne, qu'elles aient volontairement été divulguées par celle-ci, ou qu'elles aient été obtenues sur la base de son comportement en ligne passé. Les systèmes d'IA ont tendance à suggérer aux utilisatrices et utilisateurs des informations qui correspondent à leurs préférences de contenu ; les users évoluent alors principalement à l'intérieur de leur bulle de filtre. De plus, si les gens ne lisent que les articles qui leur sont suggérés par leur cercle d'amis, le danger de se retrouver dans

une chambre d'écho augmente : ici, ils ne sont en contact qu'avec des contenus qui correspondent à leurs intérêts et reflètent leur propre opinion, tandis que les positions divergentes et les informations contraires disparaissent.

Bien que les bulles de filtre numériques et les chambres d'écho attirent considérablement l'attention du public et puissent être vues comme une menace pour la démocratie, il n'existe pas à ce jour de preuve scientifiquement fondée que la diversité des opinions en ligne est plus restreinte que dans les cercles d'opinion traditionnels de personnes partageant les mêmes idées. Cependant, l'évolution décrite ci-dessus n'en est qu'à ses débuts et pourrait très bien devenir problématique pour la formation de l'opinion dans les sociétés démocratiques.

L'usine à trolls et ses théories de conspiration

Les reportages trompeurs dans les médias ne sont pas un phénomène nouveau, comme l'atteste le terme « canard » qui enrichit le monde des métaphores animales depuis le début du XIX^e siècle. En général, ils sont involontaires et nuisent à la réputation du journal qui les a publiés. Ce qui est nouveau en revanche, c'est la désinformation de masse intentionnelle par le biais des médias sociaux.

La diffusion de fausses nouvelles, ou infox (fake news), est d'ores et déjà largement automatique et très efficace, recourant aux robots sociaux (social bots), des programmes simples écrits pour diffuser des messages sur les médias sociaux de manière ciblée. Pendant la campagne électorale américaine opposant Hillary Clinton et Donald Trump en 2016, près d'un cinquième des tweets envoyés provenaient de robots sociaux ; parmi les comptes Twitter automatisés, la grande majorité (c'est-à-dire jusqu'à trois quarts d'entre eux) faisait la promotion de Trump.

L'IA pourrait à l'avenir être capable d'imiter ce qu'on appelle des trolls. Il s'agit d'utilisatrices et d'utilisateurs qui, par le biais de blogs et de forums au contenu émotionnel souvent enrichi de fausses informations, tentent de provoquer des réactions tout aussi violentes. Avec cette stratégie, les trolls, souvent employés dans de véritables « usines », visent à faire connaître certains groupes ou même à répandre des théories de conspiration. Les robots sociaux ainsi que les trolls se réfèrent souvent à des sites Internet spécifiquement conçus pour diffuser de la désinformation et augmentent ainsi la crédibilité de leurs

agents numériques. Aujourd'hui déjà, les systèmes d'IA augmentent l'efficacité avec laquelle les trolls colportent leurs messages.

La démocratie en danger

Une des stratégies des individus à l'origine de fausses nouvelles est d'infiltrer les chaînes d'information établies. Aux États-Unis, on a constaté que la désinformation est souvent présentée en premier aux médias locaux, car ceux-ci ont moins de moyens pour vérifier les faits. Une fois reprises par les nouvelles locales, les infox atteignent rapidement les médias nationaux. Cependant, l'IA ne participe pas seulement à la diffusion de fausses informations, elle contribue aussi à ce que la fausse information elle-même devienne de plus en plus élaborée. L'IA peut par exemple être utilisée pour manipuler des images, des fichiers audio et même des vidéos, ou pour créer de toutes pièces des séquences totalement fictives afin de discréditer ou de rançonner les gens. Les experts appellent ces fausses vidéos générées techniquement des hypertrucages (deepfake).

Sous les régimes autoritaires qui proscrirent les reportages déplaisants et contrôlent les médias, les médias sociaux constituent parfois un dernier recours pour obtenir des informations indépendantes. En revanche, dans les démocraties qui fonctionnent, influencer l'opinion publique par la diffusion de fausses nouvelles dans les médias sociaux est considéré comme une menace sérieuse pour le système politique. D'une part, parce que le public perd confiance dans les chaînes d'information établies : selon des enquêtes récentes, dans presque tous les pays, moins de la moitié des personnes interrogées font confiance à la capacité des médias traditionnels à distinguer les faits de la fiction. D'autre part, parce que diverses analyses d'envergure ont montré que les fausses nouvelles se répandent encore mieux que les informations exactes. En Allemagne, par exemple, il a été prouvé que sur les dix nouvelles les plus diffusées au sujet d'Angela Merkel, sept doivent être classées comme infox.

Un regard sceptique sur les réseaux sociaux

Les enquêtes menées par TA-SWISS dans le cadre de l'étude ont révélé un certain scepticisme à l'égard des médias sociaux : la majorité des gens craint que les réseaux sociaux ne nuisent à la qualité du contenu journalistique. La simple diffusion de nouvelles en ligne est, quant à elle, jugée un peu moins sévère-

ment. La personnalisation du contenu trouve ses partisans parmi les spécialistes ; il semblerait que le facteur décisif soit que les users puissent eux-mêmes décider si le flux de messages qui leur est destiné doit être personnalisé, comment et dans quelle mesure.

Les infox en revanche sont considérées comme un danger sérieux par les personnes interrogées. Seule une minorité d'entre elles croit que le progrès technologique permettra de mieux démasquer la désinformation. Au contraire, la majorité craint qu'une technologie plus élaborée ne produise à son tour des fausses nouvelles toujours plus sophistiquées. Tou-

tefois, les résultats de l'enquête ont également révélé que le droit à la liberté d'expression doit être pris en compte dans la lutte contre la désinformation. Le fait que les autorités mettent en place des contrôles de contenu spécifiques est jugé d'un œil critique. D'autre part, les personnes interrogées saluent le fait que les opérateurs de plateforme suppriment les fausses nouvelles et bloquent les individus qui en sont à l'origine. Lorsque la vérification des faits est automatisée, c'est-à-dire réalisée à l'aide de l'IA, elle est considérée comme plus utile que lorsqu'elle fait l'objet de procédures de contrôle manuelles.

L'IA dans le monde professionnel : quand les machines font le travail d'un cerveau

Pendant longtemps, on a pensé que seules les tâches simples et répétitives risquaient d'être victimes de la rationalisation. Depuis que l'IA est apparue dans des secteurs professionnels considérés comme exigeants et qui requièrent une bonne formation, la crainte de la perte d'emplois a également atteint les bureaux.

Au tournant du millénaire, différentes rubriques de journaux ont divertis leurs lecteurs avec des traductions de textes courts produits par des machines. Les programmes informatiques butaient sur les métaphores les plus simples et fournissaient des formulations fantaisistes et souvent comiques. Jusqu'à récemment, la traduction était considérée comme une tâche qui ne pouvait pas être effectuée par une machine. Pourtant, il existe aujourd'hui des logiciels tels que DeepL qui utilisent des réseaux neuronaux artificiels et sont en mesure de fournir des traductions exploitables, alors que d'autres programmes sont déjà capables de générer par eux-mêmes de simples textes journalistiques.

Le temps où l'automatisation concernait seulement les tâches simples de routine est révolu : l'IA est utilisée dans le domaine médical, où elle fournit des résultats plus précis que la plupart des professionnels de la santé dans l'évaluation des images par ultrasons, résonance magnétique ou rayons X, ou dans le domaine juridique, où elle aide les avocats à rechercher des clauses spécifiques dans des milliers de pages de contrats.

Un certain nombre d'études fournissent des chiffres sur le nombre d'emplois menacés – ou autrement dit, selon le point de vue, sur le potentiel de rationalisation. L'éventail des estimations de pertes d'emploi est large, allant de 6 à 57 % selon les hypothèses de base et la méthode utilisée. Les experts s'accordent à dire que, contrairement aux précédents bouleversements économiques liés à l'innovation, il sera vraisemblablement difficile de compenser la perte d'emploi causée par l'IA par de nouveaux produits et services.

Une carrière sous le signe de l'IA

Néanmoins, l'impact de l'IA se ressent non seulement sur la quantité mais aussi sur la qualité du travail. Par exemple, les responsables des ressources humaines utilisent de plus en plus les systèmes d'IA pour présélectionner les dossiers de candidature lors du recrutement du personnel. Les candidates et candidats doivent d'abord remplir des questionnaires standardisés en ligne. Par la suite, les responsables des ressources humaines continuent de travailler sur la base des dossiers présélectionnés par l'IA. Cela n'est pas forcément négatif pour les candidates et candidats : lorsque c'est une personne qui examine les documents soumis dès le début de la procédure, des critères de décision non objectifs tels que l'apparence photogénique peuvent se voir accorder un poids excessif. Au contraire, l'utilisation d'un système d'IA offre la possibilité de dévoiler les pratiques discriminatoires existantes et de les modifier.

Les grandes entreprises utilisent de plus en plus l'IA pour établir les profils de compétences de leur personnel et pour proposer des formations sur mesure. Cela peut poser problème dans la mesure où cela revient à limiter l'autonomie de l'individu : ce n'est pas ce dernier qui décide des mesures de qualification à prendre, mais c'est un système qui essaie d'optimiser les ressources humaines (et financières) disponibles. Même si ce n'est pas le système qui prend la décision finale, ses conseils peuvent développer une dynamique propre dans l'entreprise. Il devient difficile pour une employée ou un employé de justifier pourquoi elle ou il ne veut pas suivre une recommandation. De manière générale, la protection des données personnelles pose un problème fondamental dans le cadre de l'utilisation de l'IA dans le monde du travail.

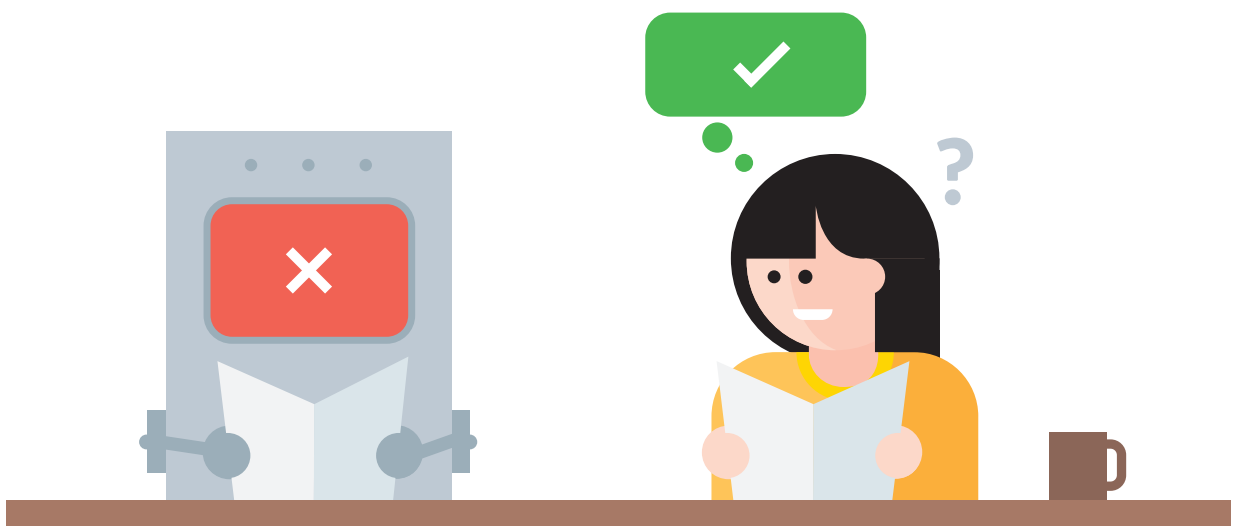
Travailler comme une machine

À l'avenir, l'IA n'entraînera pas nécessairement la perte de postes de travail entiers, mais plutôt de certaines activités individuelles. Le salariat devra apprendre à coopérer avec des machines « intelligentes » et à faire face à l'accélération du rythme et à l'intensification du travail. Les spécialistes et les institutions telles que l'Organisation internationale du travail (OIT) soulignent la nécessité de placer les personnes au centre de l'organisation du travail afin que, lorsqu'elles utilisent des systèmes d'IA, elles ne soient pas moins bien loties qu'auparavant. Il sera également important, surtout en cas de comportement inexplicable de l'IA, que les employées et employés soient en mesure d'évaluer la plausibilité des résultats et d'éviter les erreurs grossières.

L'influence de l'IA sur l'organisation du travail est difficile à distinguer des effets de la numérisation en général. Comme l'a montré l'étude de TA-SWISS sur le travail flexible, la numérisation peut contribuer à accroître la liberté personnelle et l'autonomie temporelle. Cependant, elle remet aussi en question le système actuel de sécurité sociale. Il existe un grand nombre de plateformes sur Internet qui fractionnent les grosses commandes et les envoient en portions à ce qu'on appelle des clickworkers. Cela permet aux étudiantes et étudiants de gagner de l'argent de poche supplémentaire le soir – mais pour les professionnels de la création indépendants qui utilisent ce type de plateforme pour acquérir leurs mandats sur le marché international, la pression et l'incertitude existentielle augmentent.

Des mesures contre la division du marché du travail sont requises

De l'avis des spécialistes interrogés dans le cadre de l'étude TA-SWISS, il est probable que l'IA accélère la division du marché du travail. L'écart entre le personnel hautement qualifié avec des salaires correspondants et le personnel peu qualifié à faible revenu devrait se creuser. Les expertes et experts prévoient également un contrôle accru du personnel et une instabilité croissante dans les relations de travail. En revanche, il est jugé peu probable que l'IA contribue à réduire les heures et la charge de travail. Dans l'ensemble, la majorité des expertes et experts estime que ce sont les grandes entreprises qui tireront le plus profit de ces changements dans le monde du travail.



Les spécialistes interrogés considèrent que la formation continue est une mesure particulièrement efficace contre les effets de la polarisation. La question des précautions à prendre face aux divisions croissantes dans le monde professionnel semble polariser le monde du travail tout autant que l'IA. Par exemple, une majorité relativement faible des personnes interrogées est favorable à l'atténuation des inégalités sociales par le développement du système d'assurances sociales ; une majorité tout aussi faible est opposée à la réduction du temps de travail pour répartir la diminution du nombre d'emplois entre un plus grand nombre de personnes. L'introduction d'un revenu de base inconditionnel est, quant à elle, soutenue par une majorité relative. En revanche, les auteurs de l'étude eux-mêmes considèrent qu'une répartition à la fois meilleure et plus flexible du travail et des revenus est la solution à privilégier à court comme à moyen terme. Il faudrait également promouvoir de nouveaux types d'activité

professionnelle comme des startups créatives et développer des formes d'imposition innovantes afin de continuer à financer le budget public et le système d'assurances sociales.

De l'avis des personnes interrogées, il est également nécessaire de protéger les employés et employées contre un contrôle excessif de la part de leurs employeurs. Des mesures contre la surveillance sur le lieu de travail et une meilleure protection des données en général sont considérées comme souhaitables et efficaces, le personnel devant également être protégé dans les relations de travail qui s'étendent au-delà des frontières nationales. Néanmoins, divers spécialistes soulignent également le potentiel positif de l'IA, à condition qu'elle soit utilisée pour automatiser le travail monotone et accroître la productivité et la compétitivité. Si les avantages réalisables sont répartis équitablement, c'est toute la société qui en bénéficiera.

L'IA dans la formation et la recherche : l'intelligence humaine assistée

L'IA est un produit de la recherche – et elle contribue à son tour à faire avancer la science. À l'école, les systèmes d'IA peuvent servir à apporter une aide individualisée aux élèves.

Le soutien scolaire individuel des élèves est déjà à l'agenda des politiques éducatives. L'IA laisse entrevoir des possibilités de répondre à cette préoccupation par des moyens techniques.

À chaque enfant son super professeur

De nombreuses startups ont déjà développé des outils pour aider les élèves à atteindre un objectif pédagogique. Par exemple, en utilisant des chatbots : il s'agit de programmes qui réagissent automatiquement aux questions et aux messages des apprenantes et apprenants et leur donnent des réponses préprogrammées basées sur des mots clés et des règles, ou générées individuellement par le système d'IA.

D'autres programmes guident les jeunes pas à pas au moyen de tutoriels qui leur enseignent ce qu'ils souhaitent savoir en mathématiques, en sciences

ou en physique. Ces systèmes d'IA s'appuient sur les profils de compétences et d'intérêts fournis au départ pour identifier les forces, les faiblesses et les connaissances antérieures de l'enfant ; ils analysent ensuite son mode d'apprentissage et lui suggèrent l'étape suivante la plus appropriée.

Une autre approche consiste à faire en sorte que le système d'IA surveille le processus d'écriture des apprenantes et apprenants et leur fournisse des commentaires en continu. Il existe également des systèmes d'IA qui organisent l'apprentissage en groupes en réunissant des élèves ayant des profils correspondants ou en mettant en contact des jeunes qui échouent en mathématiques avec des tuteurs de mathématiques d'autres classes, d'autres écoles ou même d'autres pays.

L'IA assiste également les enseignantes et enseignants, notamment dans le cadre de l'évaluation des performances de leurs élèves, mais aussi pour les soutenir individuellement. Grâce à l'utilisation de l'IA, le corps enseignant peut gagner plusieurs heures par semaine ne serait-ce que pour les tâches administratives.

À la recherche des talents de demain

Si les élèves effectuent de plus en plus de tâches en ligne en dialoguant avec un système d'IA, un flot de données personnelles se déverse sur les serveurs du fournisseur. Celui-ci en apprend non seulement beaucoup sur les stratégies d'apprentissage et les compétences des jeunes, mais aussi sur leurs préférences et leurs loisirs. Ces informations sont intéressantes, car elles permettent d'identifier très tôt les talents potentiels. Les données sur les enfants moins performants sont également utiles, car ceux-ci peuvent être intéressés à un cours de soutien scolaire proposé par la même entreprise.

Si une école en Suisse utilise des logiciels d'IA de fournisseurs privés, une série de problèmes liés à la protection des données se pose de manière particulièrement aiguë : Qui peut conserver, posséder ou consulter quelles données personnelles pendant combien de temps, et comment s'assurer qu'elles peuvent être supprimées si nécessaire ? Est-il possible d'emporter ses propres données en cas de changement d'école ? Et les autres écoles sont-elles autorisées à consulter les données lorsqu'elles admettent de nouveaux élèves ? Comment peut-on s'assurer que les données sur les examens et l'apprentissage ne fausseront pas la concurrence pour les emplois à l'avenir si certains élèves peuvent présenter des résultats favorables, tandis que d'autres ne disposent d'aucune donnée ou seulement de données défavorables ? Afin d'éliminer les principaux risques liés au traitement problématique des données personnelles, la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP) a lancé la plateforme edulog.ch. Elle permet aux élèves suisses de se connecter de manière anonyme aux programmes de formation. Les cantons participent à titre facultatif et il reste à savoir si l'offre sera utilisée dans la pratique.

Influence croissante des entreprises privées sur la formation

La majorité des spécialistes interrogés par TA-SWISS dans le cadre de l'étude salue la possibilité d'adapter les contenus pédagogiques à chaque élève parce que cette individualisation permet de répondre aux problèmes spécifiques de chaque apprenant et apprenante. Les expertes et experts interrogés craignent toutefois que les entreprises privées n'exercent de plus en plus d'influence sur les écoles, qui pourraient de ce fait devenir dépendantes d'un seul fournisseur.

Toutefois, interdire l'utilisation des systèmes privés d'IA n'est pas considéré comme une option viable. Il faudrait plutôt négocier avec les fournisseurs des accords précis et contraignants, comprenant des règles claires sur la protection des données. Quant à savoir si le développement de systèmes d'IA pour les écoles devrait être laissé aux autorités, l'avis des spécialistes n'est pas unanime ; la coopération entre administration publique et prestataires privés n'en reste pas moins hautement souhaitable. En majorité, ils recommandent également que les autorités procèdent à un contrôle des applications scolaires de l'IA.

L'IA fait l'objet de recherches – et fait avancer la recherche

En chiffres absolus, c'est en Chine et aux États-Unis que l'on compte le plus grand nombre de publications sur l'IA. Mais les travaux de recherche en Suisse sont également remarquables : par rapport au nombre d'habitants, la quantité d'articles scientifiques sur l'IA publiés ici est plus élevée qu'en Allemagne ou en France ; si l'on considère l'impact des citations, la Suisse se situe même parmi les premiers pays au niveau mondial.

L'IA est utilisée dans les disciplines scientifiques les plus diverses, en particulier lorsqu'il faut analyser d'énormes volumes de données, comme par exemple en science des matériaux, mais aussi en biochimie et en astrophysique. Elle est également utile dans l'enseignement, par exemple lorsqu'il s'agit de déceler des fraudes dans les articles scientifiques. Les outils utilisant l'IA détectent non seulement les passages contenant des similitudes dans les textes plagiés mais aussi les fausses statistiques. L'IA est également capable de repérer un changement de style d'écriture, ce qui peut révéler le recours à un prête-plume.

Développer les compétences clés et assurer la transparence

L'IA devient de plus en plus importante pour la recherche ; elle est même souvent la condition essentielle de certaines avancées majeures. La majorité des expertes et experts interrogés estime donc que, pour la formation continue et l'utilisation de l'IA dans la recherche, les institutions scientifiques devraient mettre en place des relais qui pourraient également servir de pôle de réflexion interdisciplinaire pour ces technologies.

Selon la majorité des personnes interrogées, les données générées par l'IA restent la propriété de l'institution de recherche concernée ou doivent être mises à la disposition du public en format libre accès. Enfin, les connaissances générées par les

systèmes d'IA doivent être aussi transparentes et reproductibles que possible. Pour répondre à cette exigence, il est nécessaire de fixer des directives appropriées en matière de transparence.

L'IA dans l'administration : des compétences souveraines pour l'intelligence artificielle ?

Les autorités travaillent généralement avec des formulaires et suivent des procédures standardisées. Ce sont des conditions idéales pour utiliser l'IA. L'accroissement de l'efficacité administrative est toutefois contrebalancé par le risque de limitation des droits fondamentaux des citoyennes et citoyens.

En septembre 2018, un conseiller national suisse du Parti socialiste a soumis au Conseil fédéral une motion le chargeant de créer un centre de compétence dans l'administration. Celui-ci avait pour but de favoriser l'utilisation de l'IA et de l'apprentissage machine afin d'accroître l'efficacité des autorités. Le Conseil fédéral a recommandé le rejet de la motion.

Lutte contre la fraude et la criminalité

L'IA a pénétré jusque dans les bureaux des fonctionnaires. Par exemple, les autorités fiscales australiennes utilisent l'IA pour recouvrer automatiquement les créances de l'État. Le système a été critiqué pour avoir tiré des conclusions partiellement erronées et pour avoir laissé aux contribuables le soin de se défendre contre une imposition excessive. Dans les bureaux des impôts allemands, l'IA est utilisée pour détecter les fraudes dans les déclarations d'impôts.

Bien que l'utilisation de l'IA dans la procédure d'imposition ne soit pas (encore) d'actualité en Suisse, la numérisation progresse également dans ce domaine : l'enregistrement et le traitement numérique des droits de douane progresse, et le prélèvement de diverses taxes de consommation, comme celles sur l'alcool, les huiles minérales et le tabac, pourraient également s'automatiser à l'avenir. L'annexe au projet de la nouvelle loi sur la protection des données adopté en septembre 2015 édicte les bases juridiques à cet égard.

La police utilise quant à elle l'IA pour estimer dans quel quartier un crime pourrait se produire dans un avenir proche. Le but de ces prévisions est de permettre aux forces d'intervention de patrouiller plus étroitement là où la probabilité d'une effraction ou d'un vol est particulièrement élevée. La « prévision policière » est particulièrement courante dans certaines grandes villes américaines. Une autre approche se concentre plutôt sur les personnes et leur réseau de relations que sur l'espace : à l'aide des données des contacts sociaux dans des cercles spécifiques, le système d'IA calcule la probabilité qu'une personne soit membre d'une organisation criminelle.

D'autre part, l'IA aide à évaluer le risque de récidive des délinquants. Aux États-Unis, on utilise un système d'IA qui calcule la probabilité qu'un contrevenant commette une nouvelle infraction en fonction de 137 critères. Le système a fait l'objet de critiques parce qu'il attribuait systématiquement un risque de récidive plus élevé aux délinquants de couleur qu'aux délinquants blancs. Un logiciel est également utilisé en Suisse pour l'évaluation criminalistique – en particulier pour déterminer si une personne donnée peut bénéficier d'une peine assouplie. Mais cet examen ne s'appuie pas sur des technologies d'IA très élaborées.

Les droits fondamentaux l'emportent sur l'arbitraire de l'État

Dans le cadre de leurs tâches souveraines, les autorités publiques traitent un grand nombre de données à caractère personnel. L'utilisation de systèmes d'IA entraînerait ici un gain considérable d'efficacité : à l'avenir, les documents et les dossiers pourraient être créés automatiquement, tandis que des assistants de bureau numériques vocaux répondraient aux demandes de renseignements courantes de la population.

Mais, par ailleurs, les autorités doivent faire preuve d'un soin particulier dans toutes leurs actions et ne doivent pas les subordonner uniquement à l'efficacité. Les droits fondamentaux garantis à l'individu face à l'État stipulent que chaque individu doit être protégé contre tout acte qui porterait atteinte aux positions protégées par les droits fondamentaux.

La protection des données personnelles est une exigence fondamentale à cet égard ; ces données doivent être stockées en toute sécurité et protégées contre toute utilisation abusive. Chacun doit également pouvoir partir du principe que les données recueillies et traitées à son sujet sont exactes, complètes et équilibrées. En fait, de nombreuses personnes ne savent pas quelles données les concernant sont stockées et utilisées dans le cadre d'un traitement utilisant l'IA ; elles ne sont donc guère en mesure de suivre ni de corriger les erreurs éventuelles dans l'ensemble des données recueillies sur elles.

Au vu des distorsions et des déséquilibres qui peuvent être associés à un traitement des données utilisant l'IA, les procédures automatisées sont particulièrement délicates car elles pourraient entraîner une discrimination à l'égard de certains groupes et individus. La révision de la loi sur la protection des données stipule ici que les personnes concernées doivent être informées en cas de décisions entièrement automatisées.

Finalement, le fait de laisser la décision à un système d'IA entièrement automatisé entre en contradiction avec l'exigence légale selon laquelle une instruction des autorités ou un jugement doit être justifié. Tout acte officiel doit être présenté

de manière transparente afin que les personnes concernées puissent, le cas échéant, s'y opposer. Toutefois, si une machine auto-apprenante basée sur l'IA prend une décision automatisée ou effectue des tâches préparatoires essentielles à cette fin, il se peut qu'il devienne impossible de retracer le raisonnement du système d'IA. De même, il deviendrait impossible de motiver la décision d'une manière qui réponde aux exigences légales.

La Suisse reste prudente

Dans le sondage réalisé dans le cadre de l'étude TA-SWISS, la majorité des spécialistes était d'avis que l'utilisation de systèmes d'IA dans l'administration publique était moins probable en Suisse que dans d'autres pays et qu'une automatisation complète des procédures officielles complexes n'était guère envisageable. En Suisse, les systèmes d'IA pourraient tout au plus être utilisés dans des contextes simples et des processus partiellement automatisés.

En revanche, la majorité des personnes interrogées estime que le risque que les autorités suisses deviennent dépendantes de fournisseurs de technologie étrangers est considérable. Selon les spécialistes, la menace d'une certaine soumission aux machines existe également, en ce sens que le personnel d'administration pourrait peiner à justifier une position différente de celle de l'IA. Enfin, les expertes et experts interrogés mettent en garde contre une fragilisation de la protection des données et contre le fait que les processus automatisés ne peuvent être contrôlés que de manière insatisfaisante.



Pas de délégation du pouvoir décisionnel à la machine

Les spécialistes interrogés s'accordent à une nette majorité : lorsqu'il s'agit d'analyses prospectives ou de la création d'un profil de personnalité (« profilage »), les systèmes d'IA ne peuvent que formuler des recommandations ; la décision elle-même doit être laissée aux êtres humains. Une faible majorité des personnes interrogées estime également que l'IA ne devrait être utilisée que pour préparer des décisions relatives à un espace ou un objet, mais pas celles qui concernent des individus.

La majorité salue le fait que l'État soit tenu de dévoiler dans quelles procédures il utilise l'IA. Il faut en outre que les autorités expliquent de manière compréhensible les grandes lignes des processus qui requièrent l'utilisation de l'IA. Et enfin, l'État doit justifier individuellement chaque décision qui se base sur l'IA.

Les machines « intelligentes » – un défi pour la législation en vigueur

Si une ingénieure se trompe en calculant la portance d'un mur ou si un traducteur commet une erreur en traduisant un texte dans une autre langue, il n'existe en principe aucun doute quant à la personne responsable des dommages. Mais que se passe-t-il lorsque les erreurs découlent de l'utilisation d'un système d'IA ? L'IA soulève de nouvelles questions juridiques et éthiques, qui ne peuvent être abordées que superficiellement ci-dessous.

Actuellement, l'ordre juridique présuppose que la décision d'un sujet est à l'origine de toute action, et que tout ce qui se passe ensuite en découle, y compris les conséquences légales. Cette notion, qui part du principe que tout dommage peut être attribué à son auteur, est remise en cause par l'IA. Le problème s'accroît à mesure que les systèmes d'IA prennent des décisions plus indépendantes et imprévisibles.

Tenir les systèmes d'IA en laisse ?

L'idée de devoir assumer la responsabilité de dommages qu'elle n'a pas directement causés est le cauchemar de toute personne possédant un chien : peu importe que le compagnon à quatre pattes ait creusé des trous dans le jardin voisin ou qu'il ait mordu la main de la factrice, c'est sa ou son propriétaire qui doit rembourser le dommage ou supporter les conséquences. La ou le propriétaire ne peut être dégagé

de sa responsabilité que s'il peut prouver qu'il a fait preuve de la diligence requise avec son animal.

Divers juristes estiment que la responsabilité du détenteur d'animaux pourrait servir de modèle pour une responsabilité spécifique des systèmes d'IA autonomes : ces derniers aussi sont parfois imprévisibles et ont le potentiel de causer des dommages. Les détracteurs de cette analogie soulignent que l'activité de l'IA qui a conduit au dommage a été déléguée à l'IA, de sorte que c'est la personne qui a utilisé un système d'IA qui se trouve à l'origine de toute la chaîne d'action.

Une approche plus ambitieuse permettrait de donner aux systèmes d'IA un statut juridique propre. Depuis 2017, la Commission européenne envisage la création d'une « personne électronique » qui permettrait de tenir les systèmes robotiques sophistiqués pour responsables des dommages qu'ils causent ; une approche similaire serait également envisageable pour les systèmes d'IA mais, d'après les spécialistes, elle serait encore clairement prématurée.

Créativité de première et de seconde main

Les méthodes mathématiques sont exclues de la brevetabilité en vertu de la législation en vigueur. La situation est moins claire en ce qui concerne la protection par brevet des programmes informatiques. Selon la pratique de l'Office européen des brevets, la brevetabilité peut être accordée si l'apprentissage machine apporte une « contribution technique » à une invention – par exemple, en utilisant des réseaux neuronaux pour classer des images et des vidéos numériques. La possibilité de breveter un système d'IA doit donc être examinée au cas par cas.

Des questions sont également ouvertes en ce qui concerne la protection éventuelle du droit d'auteur. Les programmes informatiques sont en effet considérés comme des œuvres pouvant bénéficier de la protection du droit d'auteur. Cependant, le problème se pose avec les systèmes qui apprennent d'eux-mêmes et continuent ainsi de se développer en permanence. Dans ce cas, l'objet de la protection du droit d'auteur change car il peut arriver qu'avec le temps, un système d'IA ne soit plus le même qu'au début.

Une autre difficulté se pose dans le cas des biens immatériels produits par les systèmes d'IA. Il n'est pas clair si de tels biens générés par l'IA peuvent être protégés et, dans l'affirmative, qui doit être considéré comme leur créateur du point de vue du droit de la

propriété intellectuelle : le programmeur qui a développé le système en question, la personne qui l'utilise, voire le système lui-même ? Si cette dernière option est déjà exclue du fait que les systèmes d'IA ne sont pas des personnes juridiques, l'attribution à une personne physique du droit de la propriété intellectuelle en question doit faire l'objet d'une distinction : la protection des œuvres générées de manière totalement autonome par l'IA risque déjà d'être impossible par le simple fait qu'aucun être humain n'a contribué intellectuellement à leur création, ce qui est une condition préalable à la protection d'une création dans le cadre de la loi sur le droit d'auteur. Dans la pratique, cependant, il est difficile de distinguer ces œuvres des autres productions dans lesquelles l'IA n'a servi qu'à titre d'outil auxiliaire.

En revanche, du point de vue du droit des brevets, il faut partir du principe que les inventions générées par l'IA ont un caractère fondamentalement protégeable. La personne physique devient un inventeur au sens juridique du terme car elle est la première à prendre connaissance du résultat et à le comprendre comme une solution à un problème technique.

L'équité est difficile à calculer

L'IA soulève des problèmes difficiles à résoudre non seulement sur le plan juridique mais aussi au niveau éthique. Le système américain d'IA COMPAS, qui s'appuie sur 137 critères pour évaluer la probabilité qu'un criminel récidive, en est une illustration. Le système a fait l'objet de critiques lorsqu'il a été démontré qu'il calculait à tort un risque de récidive pour les Afro-Américains presque deux fois plus élevé que pour les Blancs ; par contre, la probabilité qu'un contrevenant récidive était prédite presque deux fois moins souvent chez les Blancs que chez les Noirs. L'algorithme n'avait pourtant reçu aucune information sur l'origine ethnique des prévenus.

L'entreprise proposant le test s'est défendue contre l'accusation de racisme en affirmant qu'elle avait mis au point un test pour s'assurer que le taux de précision était le même pour les personnes à la peau foncée et pour celles à la peau claire – et que cela garantissait un logiciel équitable. Les mathématiciens se sont ensuite penchés sur la question de savoir si un test peut être équitable et inéquitable à la fois, et ont conclu que cela est effectivement possible – à savoir lorsqu'une caractéristique (en l'occurrence la récidive) se produit avec une fréquence différente dans une population en fonction du sous-groupe. Comme les Afro-Américains récidivent plus souvent que les Blancs, il ne serait pas possible de faire des

prédictions aussi précises pour les deux groupes et d'obtenir en même temps les mêmes taux d'erreur pour les deux groupes ethniques. La justice, selon les spécialistes, ne peut être établie statistiquement ou mathématiquement. Il appartient plutôt à l'éthique d'élaborer des procédures et des critères pour assurer ou au moins améliorer l'équité des algorithmes.

Lignes directrices en matière d'éthique au niveau européen

Au printemps 2019, un groupe d'expertes et d'experts en IA a publié, au nom de la Commission européenne, des lignes directrices en matière d'éthique pour une IA digne de confiance. Les directives sont destinées aux personnes qui développent l'IA, à celles qui l'utilisent et à celles qui exploitent ses résultats. Elles visent à garantir le respect des valeurs fondamentales et des réglementations applicables et à ce que toute utilisation d'IA poursuive un but éthiquement souhaitable.

Les directives soulignent la nécessité de veiller à ce que le contrôle des décisions et des procédures reste de la compétence d'un être humain, même lorsque l'IA est utilisée. Les systèmes technologiques doivent également être robustes et fournir des résultats fiables et reproductibles. Les données et la vie privée doivent être protégées, et les systèmes d'IA doivent être utilisés de manière transparente pour préserver la traçabilité des voies décisionnelles. Enfin, il est important d'éviter les distorsions inévitables et de garder à l'esprit le bien-être social et environnemental lorsque des systèmes d'IA sont utilisés.

Distorsions et manque de transparence

Du point de vue de la majorité des spécialistes interrogés, le déséquilibre des ensembles de données et le manque de transparence dans l'utilisation de l'IA présentent des risques considérables. Cependant, les expertes et experts sont généralement favorables à l'utilisation de l'IA et lui attribuent une influence positive sur les objectifs sociaux largement acceptés, comme la lutte contre la pauvreté ou le renforcement de la santé.

Il n'en reste pas moins que, de l'avis des personnes interrogées, la responsabilité des décisions en matière d'IA doit être assumée par l'être humain. Les utilisatrices et utilisateurs doivent être informés lorsqu'ils interagissent avec l'IA, et ils devraient également être capables de comprendre comment une IA est arrivée à sa décision.

Recommandations : ouvrir le champ

Les systèmes d'IA sont utilisés dans les domaines d'application les plus variés. Les recommandations en matière d'utilisation de cette technologie couvrent par conséquent un champ très vaste.

Les systèmes d'IA sont utilisés pour les tâches les plus diverses dans des domaines très différents, chacun d'entre eux soulevant également ses propres questions sur le plan juridique. L'élaboration d'une « loi sur l'IA », qui viserait à réglementer de manière uniforme l'utilisation de l'IA, n'est donc pas appropriée et doit par conséquent être rejetée.

Examiner les applications spécifiques

Qui utilise un système d'IA à quelle fin exactement, sur quelles données celui-ci se base-t-il et quelles sont les exigences légales à prendre en compte ? Ces questions doivent recevoir une réponse avant que les chances et les risques liés à l'utilisation de l'IA ne puissent être évalués. Il faut donc toujours fixer les règles d'utilisation de l'IA en fonction de chaque contexte d'utilisation. Il convient également d'identifier les problèmes concrets et les évolutions indésirables, et de les résoudre avec les précautions appropriées ; lorsque le risque n'est pas clairement établi, la recherche devrait être intensifiée pour déterminer les dangers éventuels.

Des exigences élevées à l'égard des autorités

Lorsque l'État exerce ses fonctions souveraines, il agit à partir d'une position supérieure par rapport à l'individu. Il est donc d'autant plus important que la légalité des actions de l'État puisse être vérifiée par les personnes concernées et que les systèmes d'IA soient utilisés de manière suffisamment transparente. Ainsi, si l'État utilise l'IA, il faut par principe qu'il réponde à des exigences plus strictes que les utilisatrices et utilisateurs du secteur privé.

Transparence et information assimilable

Lorsque les personnes sont affectées de manière notable par l'utilisation de l'IA, elles doivent être informées de manière simple et compréhensible du fait qu'elles interagissent avec un système d'IA. Néanmoins, il convient de souligner qu'un excès d'information peut conduire à une certaine résignation et donc à de la négligence de la part des personnes concernées. Il faut donc veiller à ce que ces personnes, ainsi que les organisations qui fournissent une assistance juridique aux individus lésés, obtiennent sur demande toutes les données qui leur permettent de déceler et d'évaluer les résultats erronés.

Un label de qualité au lieu d'une autorisation générale de mise sur le marché

Étant donné que les possibilités d'application des systèmes d'IA sont variées et variables, il ne serait pas approprié de soumettre leur utilisation à une autorisation générale de mise sur le marché. Lorsque, comme c'est le cas pour les produits médicaux, une autorisation de mise sur le marché est déjà requise, celle-ci comprend également l'examen de l'IA. Toutefois, les initiatives privées visant à créer des labels de qualité pour l'IA doivent être saluées et encouragées. Les organisations de protection des consommateurs, notamment, devraient être habilitées à mieux évaluer la qualité de ces certificats.

Voir au-delà de la technologie

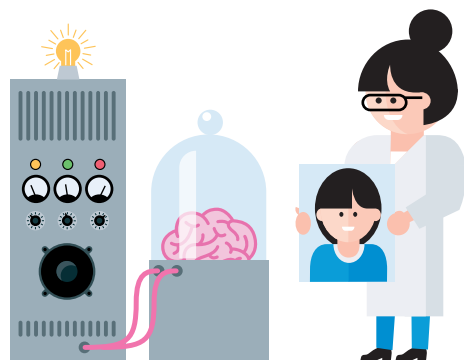
La conception et le développement de l'IA ne sont pas seulement une question de technologie, loin de là. Pour être notamment en mesure d'évaluer les risques encourus, il est essentiel que les spécialistes qui développent l'IA possèdent également une compréhension de l'éthique et des notions juridiques de base et qu'ils soient aussi disposés et capables de collaborer sur une base interdisciplinaire. Des connaissances appropriées sont également indispensables pour toute personne qui utilise les systèmes d'IA ou exploite leurs résultats.

Groupe d'accompagnement

- Prof. Dr. Jean Hennebert, Leitungsausschuss TA-SWISS, Département d'informatique de l'Université de Fribourg, Président der Begleitgruppe
- Benjamin Bosshard, Eidgenössische Kommission für Kinder- und Jugendfragen
- Sabine Brenner, Geschäftsstelle Digitale Schweiz, Bundesamt für Kommunikation (BAKOM)
- Dr. Christian Busch, Staatssekretariat für Bildung, Forschung und Innovation (SBFI)
- Dr. Christine Clavien, Institut Ethique Histoire Humanités, Université de Genève
- Daniel Egloff, Staatssekretariat für Bildung, Forschung und Innovation SBFI
- Andy Fitze, SwissCognitive – The Global AI Hub.
- Matthias Holenstein, Stiftung Risiko-Dialog
- Dr. Marjory Hunt, Fonds national suisse de la recherche scientifique (FNS)
- Manuel Kugler, Schweizerische Akademie der Technischen Wissenschaften (SATW)
- Thomas Müller, TA-SWISS Leitungsausschuss, Redaktor Schweizer Radio SRF
- Katharina Prelicz-Huber, TA-SWISS Leitungsausschuss (bis 2019), Präsidentin Gewerkschaft VPOD/SSP, Nationalrätin
- Prof. Ursula Sury, Rechtsanwältin und Professorin, Hochschule Luzern (HSLU)
- Dr. Stefan Vannoni, TA-SWISS Leitungsausschuss, cemsuisse

Gestion du projet chez TA-SWISS

- Dr. rer. soc. Elisabeth Ehrensperger, Geschäftsführerin
- Dr. Catherine Pugin, Projektleiterin



Impressum

Sans neurones mais intelligent
Synthèse de l'étude «Wenn Algorithmen für uns entscheiden:
Chancen und Risiken der künstlichen Intelligenz»
TA-SWISS, Berne 2020
TA 72A/2020

Rédaction : Lucienne Rey, TA-SWISS, Berne
Traduction : pro-verbial, Zürich
Production : Fabian Schluep, TA-SWISS, Berne
Mise en page et illustrations : Hannes Saxer, Berne
Impression : Jordi AG – Das Medienhaus, Belp

TA-SWISS – Fondation pour l'évaluation des choix technologiques

Souvent susceptibles d'avoir une influence décisive sur la qualité de vie des gens, les nouvelles technologies peuvent en même temps comporter des risques nouveaux, qu'il est parfois difficile de percevoir d'emblée. La Fondation pour l'évaluation des choix technologiques TA-SWISS s'intéresse aux avantages et aux risques potentiels des nouvelles technologies qui se développent dans les domaines « biotechnologie et médecine », « société de l'information » et « mobilité / énergie / climat ». Ses études s'adressent tant aux décideurs du monde politique et économique qu'à l'opinion publique. TA-SWISS s'attache, en outre, à favoriser par des méthodes participatives, l'échange d'informations et d'opinions entre les spécialistes du monde scientifique, économique et politique et la population. TA-SWISS se doit, dans tous ses projets sur les avantages et les risques potentiels des nouvelles technologies, de fournir des informations aussi factuelles, indépendantes et étayées que possible. Elle y parvient en mettant chaque fois sur pied un groupe d'accompagnement composé d'experts choisis de manière à ce que leurs compétences respectives couvrent ensemble la plupart des aspects du sujet à traiter.

La fondation TA-SWISS est un centre de compétence des Académies suisses des sciences.



TA-SWISS
Fondation pour l'évaluation
des choix technologiques
Brunngasse 36
CH-3011 Berne
info@ta-swiss.ch
www.ta-swiss.ch

membr des
 académies suisses
des sciences