

Vers un management responsable des services digitaux

“Chartes d’entreprise sur l’Intelligence Artificielle”

Luis-Antoine Berne, Anne Eloy-Perrot, Rose Rondelez et Elena Susini

Étudiants en Master 1 à Sciences Po - École du Management et de l’Innovation

Introduction

L’Intelligence Artificielle (IA) se retrouve au cœur de toutes les préoccupations, puisqu’elle est utilisée par un nombre croissant de sociétés. Elle résulte du travail de l’humain, de développeurs, de chercheurs. Elle fonctionne par conséquent à travers des biais, de par sa construction même. Les entreprises se retrouvent dans une position où elles doivent d’une part s’emparer de ces technologies pour rester concurrentielles, mais d’autre part apprendre à les maîtriser, pour éviter les différents biais pouvant lui porter préjudice. Par conséquent se posent peu à peu des questions d’éthique et de sécurité liées à l’utilisation de ces technologies. En réaction, des chartes de bonne conduite se mettent progressivement en place pour y répondre.

À travers la création de chartes par les entreprises et par certaines institutions et groupes comme le *High Level Experts Group on AI*¹, *Impact AI*² ou encore *Partnership on AI*³, nous avons observé une volonté de répondre à ce double défi et de mettre en place une gouvernance dédiée à l’utilisation de l’IA.

La première question à aborder est celle de la valeur juridique pour les chartes d’entreprise. Il existe différentes sources internes de droit dans les entreprises, la principale étant le règlement intérieur. Le code du travail le définit comme un document écrit par lequel l’employeur fixe exclusivement des règles en matière d’hygiène, de sécurité et de sanctions⁴.

¹ groupe d’experts en IA dont le travail a permis l’élaboration des politiques prises par la Commission Européenne et ses États membres vis-à-vis de l’IA

² association créée par Microsoft qui travaille sur les questions éthiques de l’IA

³ organisation multipartite qui rassemble des universitaires, des chercheurs, des organisations de la société civile, des entreprises, etc.

⁴ article L1321-1 du Code du travail

L'établissement du règlement intérieur est soumis à la consultation des représentants du personnel et à un contrôle de l'inspection du travail. Sans ces formalités, le règlement ne peut pas avoir d'effet. Contrairement au règlement intérieur, dont la mise en place est obligatoire pour les entreprises de plus de 50 salariés, les chartes éthiques ne résultent d'aucune obligation légale en France. Leur contenu est souvent très hétérogène : elles regroupent en effet des règles prévues par d'autres textes - lois, conventions collectives, règlement intérieur, contrats de travail - et des dispositions nouvelles. Afin d'en apprécier l'efficacité, il peut être opportun de s'interroger sur leur opposabilité aux employés et aux tiers ou, en d'autres termes, sur les effets d'un non-respect de leurs dispositions.

Premièrement, si la charte est assimilée à un élément du règlement intérieur - qu'elle a donc été signée par le salarié et qu'elle est annexée au règlement - alors les règles qu'elle contient sont opposables aux salariés ; ils pourront donc être sanctionnés en cas de non-respect de ces règles. En ce qui concerne l'opposabilité aux tiers de l'entreprise, lorsque la charte acquiert un caractère contractuel, c'est-à-dire lorsqu'elle a été un élément déterminant pour le cocontractant, alors sa violation peut engager la responsabilité de l'entreprise et la contraindre à payer des dommages-intérêts. En réalité, les dispositions nouvelles présentes dans les chartes revêtent souvent une faible valeur juridique car elles n'engagent jamais explicitement l'employeur : plutôt que d'être des règles de droit, elles sont des règles de normalisation ou règles de gestion, de simples engagements moraux, trop flous pour engager l'entreprise et donc efficaces afin de la protéger en cas de non-conformité à ses dispositions.

Là est donc tout le défi à relever : il ne s'agit pas de faire de l'Ethical-washing mais de véritablement veiller à la mise en œuvre de mesures concrètes pour éviter les biais liés à l'utilisation de l'IA. Depuis les années 2010 et plus particulièrement à partir de 2016, plusieurs événements et dates clés ont encouragé ce mouvement de responsabilisation des entreprises vis-à-vis de l'IA.

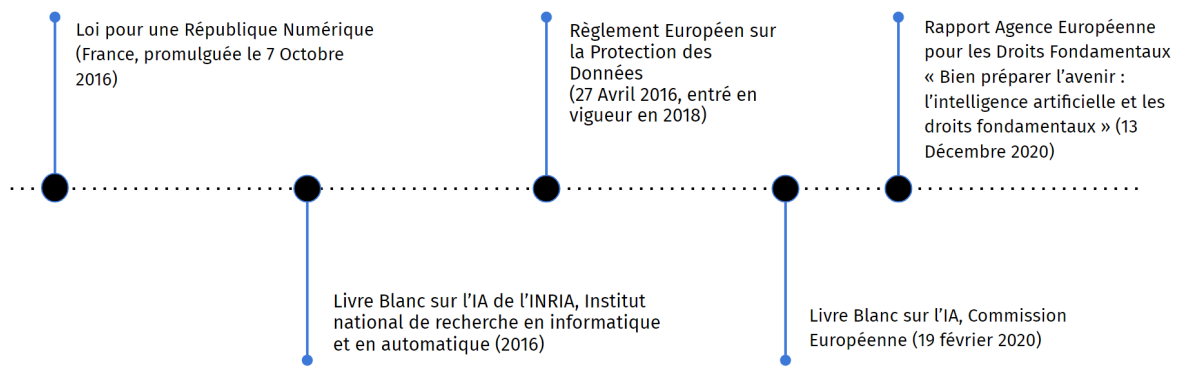


Figure 1 : Frise chronologique de quelques dates clés du mouvement de responsabilisation des entreprises sur le sujet de l'Intelligence Artificielle

À l'échelle nationale, la loi pour une République Numérique (promulguée en France le 7 Octobre 2016) est un premier jalon qui se veut rendre plus accessible les savoirs liés à l'économie de la donnée. Cette loi a aussi pour but d'œuvrer à la protection des données personnelles et de créer un cadre de confiance qui garantit le droit des utilisateurs.

Le livre Blanc sur l'IA de l'INRIA (Institut national de recherche en informatique et en automatique), publié en 2016, est lui aussi un document important qui aborde avec un œil critique les grandes tendances et les principaux défis de l'intelligence artificielle et décrit la manière dont ses équipes conduisent des recherches scientifiques, développent des logiciels et œuvrent pour le transfert technologique afin de relever ces défis.

Adopté le 14 Avril 2016 et applicable à partir du 25 Mai 2018, le Règlement Général sur la Protection des Données (RGPD) du Parlement Européen est, quant à lui, le texte de référence en matière de protection des données à caractère personnel. Cette mesure est particulièrement essentielle pour une plus grande responsabilité des services digitaux puisqu'elle permet, dans un premier temps, d'accroître la protection des individus concernés par un traitement de leurs données à caractère personnel et, dans un second temps, de veiller à la responsabilisation des acteurs de ce traitement en renforçant le pouvoir des autorités de contrôle.

On peut également citer parmi ces dates clés le rapport "Pour une intelligence artificielle maîtrisée, utile et démystifiée" qui concentre 15 propositions adoptées par l'Office

parlementaire d'évaluation des choix scientifiques et technologiques (OPECST) le 14 Mars 2017. Enfin, il convient d'évoquer le Livre Blanc sur l'IA de la Commission Européenne (19 février 2020) ainsi que le rapport de l'Agence Européenne pour les Droits Fondamentaux (13 Décembre 2020) qui souligne les risques liés à l'IA.

Certaines remarques peuvent être faites concernant le livre blanc sur l'IA de 2020 et les différents rapports adoptés par la Commission Européenne la même année. Premièrement, l'IA est déjà soumise à la législation européenne sur les droits fondamentaux, la protection des consommateurs, la responsabilité des produits et la protection des données⁵. Toutefois, l'opacité des algorithmes peut rendre l'application de ces législations peu efficace, d'où l'importance de réguler l'IA avec un régime spécifique. Ainsi, le rapport du député Iban García del Blanco propose par exemple la création d'un cadre pour les aspects éthiques de l'intelligence artificielle ainsi qu'une liste de principes éthiques que les développeurs se doivent de respecter.

Une des principales difficultés dans l'élaboration d'une future législation consistera à définir ce qu'est l'IA, de manière assez souple pour tenir compte des progrès techniques à venir, tout en étant suffisamment précis pour assurer la sécurité juridique nécessaire. Pour la Commission européenne, les obligations du nouveau cadre réglementaire sur l'IA s'appliqueraient uniquement aux solutions d'IA identifiées comme étant "à haut risque" conformément à deux critères cumulatifs : premièrement, l'IA devra être utilisée dans un secteur où l'on peut s'attendre à des risques importants (i.e. santé, transports, énergie); de plus, son utilisation devra être en mesure de produire des effets sur les droits des individus ou des entreprises et comporter des risques de blessures, décès ou dommages matériels ou immatériels importants.

Une autre question fondamentale concerne le régime de responsabilité à appliquer à l'IA. En effet, une IA peut provoquer des dommages corporels ou matériels (i.e. accident provoqué par une voiture autonome) mais aussi moraux lorsqu'elle porte atteinte à l'honneur ou à la pudeur des personnes sur les réseaux sociaux. Qui de l'humain ou du robot est responsable ? Il n'existe pas aujourd'hui de régime de responsabilité civile propre à l'intelligence artificielle et donc aucune réponse claire à cette question. Cependant, le droit de la responsabilité civile a pour but d'indemniser les victimes de dommages et les robots n'ont à ce jour pas de personnalité juridique et donc pas de patrimoine: de ce fait, il est plus

⁵ notamment via le RGPD

efficace de se tourner vers l'humain, c'est-à-dire vers la personne la plus solvable. La Commission européenne préconise d'ailleurs que l'obligation soit supportée par l'acteur de la chaîne le plus à même de faire face au risque. Ainsi, le rapport du député Axel Voss propose une responsabilité stricte pour les IA à haut risque ainsi qu'une assurance obligatoire pour les opérateurs du système. Les systèmes d'IA qui ne figurent pas dans la liste des systèmes d'IA à haut risque et qui n'ont pas provoqué de dommages ou de préjudices graves devraient donc rester en principe soumis au régime de responsabilité pour faute. Il sera donc nécessaire de comprendre et de pouvoir retracer les décisions prises par l'IA afin de caractériser sa faute et le lien de causalité entre la faute et le dommage.

Comme nous le soulignerons au cours de ce papier, nous avons tout de même constaté que les mesures évoquées plus haut permettent l'adoption progressive de rapports de plus en plus précis sur la question, en Europe principalement, mais aussi à l'internationale. En parallèle on observe aussi une mise en place de chartes par des entreprises ou associations telles que Arborus, La MAIF, Le Crédit Agricole, Google, Microsoft ou encore Sony qui sont des cas que nous avons étudiés au cours de nos recherches. En ce sens, nous en sommes venus à nous poser la question suivante : ***Dans quelle mesure les chartes d'entreprises sur l'IA traduisent-elles les grands principes éthiques pour une technologie responsable ?***

Concernant notre analyse des différentes chartes d'entreprises sur l'Intelligence Artificielle, il est temps de vous présenter notre carte heuristique (voir en annexe). Celle-ci a été organisée autour des quatre grands principes d'éthique : le principe de justice, le principe d'autonomie, le principe de bienveillance et le principe de non-malveillance. Sur ces différentes branches ont, en premier lieu, été triés les principes de grandes institutions concernant l'IA. Nous nous sommes particulièrement intéressés à Impact AI, le High Level Expert Group (HLEG) missionné par l'Union Européenne et Partnership on AI (PAI), décrits plus haut. Certains principes de ces instances se sont avérés être fortement similaires tandis que d'autres étaient beaucoup moins partagés. Par la suite, nous avons fait le lien entre ces principes et ceux de chartes particulières d'entreprises. Pour cette étude nous avons notamment étudié les principes de Microsoft, SAP, Sony, Google, MAIF, Crédit Agricole et Arborus (association de différentes entreprises comme Orange, EDF, L'Oréal et Danone).

Tout d'abord, a été associé au principe de justice un principe similaire au HLEG et au PAI : **Diversité, non-discrimination et équité**. Pour citer le HLEG dans *Ethics Guidelines*

*for Trustworthy AI*⁶, “les systèmes d’IA devraient prendre en compte tout l’éventail des capacités, aptitudes et besoins humains, et leur accessibilité devraient être garantie”. Cette revendication est particulièrement bien illustrée par l’article de la charte de SAP “We are driven by our values” qui évoque les droits de l’homme, les principes de l’ONU, les lois et les normes internationales comme des centres d’attention particuliers de l’entreprise au regard des usages de l’IA. Un second principe institutionnel attaché au principe de justice est le principe d’**Équité** évoqué par Impact AI. Bien que proche du précédent principe, nous avons souhaité le séparer car il comporte une précision particulièrement intéressante sur les biais et les injustices que peut engendrer l’IA. Le collectif Arborus donne un exemple de ce principe dans son article “Veiller à la qualité des données”, qui insiste sur les enjeux importants des bases de données. Le dernier principe de justice est le principe de **Responsabilité** évoqué conjointement par le HLEG et Impact AI. (Il faut toutefois préciser que cette similarité n’est pas un hasard puisque Impact AI a travaillé à l’élaboration de *Ethics Guidelines for Trustworthy AI* avant de publier son rapport pour une *IA digne de confiance*⁷.) Selon ce principe, “il convient de mettre en place des mécanismes pour garantir la responsabilité à l’égard des systèmes d’IA et de leurs résultats, et de les soumettre à une obligation de rendre des comptes”. Peu représenté dans les chartes d’entreprises, on retrouve cependant une définition de ce principe chez Microsoft (accompagnée d’une vidéo explicative) : “People should be accountable for AI systems”⁸.

Le deuxième grand principe éthique étudié est le principe d’autonomie. Celui-ci comprend une première notion : **Action et contrôle humain** (énoncée par le HLEG), identique au principe de **Dignité** d’Impact AI. Il consiste à penser une intelligence au service de l’humain et ne restreignant pas son autonomie. Dans les chartes d’entreprise se trouvent différentes modalités de contrôle de l’homme sur la technique. On peut prendre l’exemple du principe “Piloter” de la MAIF: “L’Homme doit toujours conserver la maîtrise de la technique. Nous garantissons que les technologies et les algorithmes que nous utilisons sont placés sous le contrôle de l’Homme.”⁹ Partnership on AI propose ensuite deux autres principes d’autonomie. Le premier, **Address concerns**, se traduit par un principe de règlement des problèmes engendrés par l’IA. Il est illustré dans la charte de Google par le principe “Be

⁶ <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

⁷

https://www.impact-ai.fr/wp-content/uploads/2020/05/IMPACTAI_Guide_IA_dignedeconiance_WEB.pdf

⁸ <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6>

⁹

<https://entreprise.maif.fr/files/live/sites/entreprise-Maif/files/pdf/nos-chartes/charte-numerique-maif.pdf>

accountable to people”, beaucoup plus axé sur une notion de contrôle humain et de feedback que sur une réelle préoccupation de responsabilité face aux conséquences de l’IA. Le Principe de PAI **Research and dialogue** promeut, quant à lui, une participation de la recherche dans le sens de l’autonomie humaine.

Le principe suivant, le principe de non-malveillance, est largement représenté dans les chartes d’entreprise, notamment parce qu’il abrite la robustesse technique et le respect de la vie privée par les technologies d’IA. Le principe **Robustesse technique et sécurité** est fortement partagé entre le HLEG, Impact AI et PAI. Il encourage la création d’algorithmes sûrs, une gestion des erreurs et une certaine sécurité des données. La grande majorité des entreprises étudiées y font référence. Un deuxième principe largement représenté est celui de **Respect de la vie privée et gouvernance des données** du HLEG et de PAI. Il est bien évidemment lié aux récentes évolutions réglementaires sur les données comme le RGPD. Le principe **Gouvernance des données** d’Impact AI a été séparé du précédent pour appuyer le fait qu’il comprend un spectre bien plus large d’action pour garantir une qualité, une intégrité et une pertinence des données. Il est en ce sens intéressant de mettre en avant le principe de “Sensibilisation aux biais” d’Arborus : “sensibiliser les prescripteurs des solutions à base d’IA (RH, finances, relations clients, marketing ...), aux risques de biais et stéréotypes pouvant générer des discriminations et intégrer dans les cahiers des charges des points de contrôle et d’évaluation itérative.”¹⁰

Le principe de bienveillance est lui aussi fortement présent dans les chartes d’entreprise. Il comprend, tout d’abord, le principe de **Transparence**. C’est, en effet, un des grands enjeux de l’IA, qui est encore bien souvent une boîte noire, tant pour les concepteurs que pour les utilisateurs. Le principe de transparence est cependant difficile à formaliser en raison des limites scientifiques et des enjeux d’entreprise comme la concurrence (fournir le code des algorithmes d’IA a peu de sens dans une stratégie concurrentielle). C’est pourquoi les principes **Understandability** et **Engage**, respectivement axés sur la compréhension du fonctionnement et sur la communication sur l’IA, semblent plus adaptés et représentatifs des possibles leviers des entreprises en matière de transparence. En ce sens, le principe “We strive for transparency and integrity in all that we do” de SAP couvre un champ très large correspondant à ces trois principes : “Their [our systems’] input, capabilities, intended purpose, and limitations will be communicated clearly to our customers, and we provide

¹⁰ <https://charteia.arborus.org/>

means for oversight and control by customers and users. They are, and will always remain, in control of the deployment of our products.”¹¹ L’entreprise allemande dévoile ici sa volonté de communication sur les différents enjeux de l’IA ainsi que son objectif d’interprétabilité et même de contrôle par les usagers de ces solutions complexes. Un principe important, évoqué par le HLEG et Impact AI et celui de **Bien-être social et environnemental**. Celui-ci promeut un impact positif de l’utilisation des algorithmes d’IA : “les systèmes d’IA devraient être utilisés pour soutenir des évolutions sociales positives et renforcer la durabilité et la responsabilité écologique”. Nous avons également rattaché à ce principe la volonté d’excellence scientifique de Google, connexe avec le principe d’autonomie **Research and dialogue**. Il illustre une envie de faire de nouvelles avancées dans la science bénéfiques à l’humanité. Le principe de bien-être social et environnemental est par ailleurs lié au principe nommé **Empower people** de PAI. Ce dernier comprend par exemple la décision de connexion aux challenges sociétaux de SAP (“We engage with the wider societal challenges of AI”) ou encore le vœu de bénéfice social de Google (“Be socially beneficial”). Le dernier principe, énoncé par PAI, est nommé **Cooperation and trust**. Il appelle à la coopération des entreprises et à une confiance mutuelle afin de parvenir à affronter les challenges cités précédemment. Ce principe étant un engagement fort pouvant s’opposer à des volontés concurrentielles, nous n’en avons pas trouvé d’illustration dans les chartes d’entreprises étudiées.

La carte heuristique ainsi présentée nous permet de tirer plusieurs conclusions que nous avons regroupées en deux parties : d’abord sur les enjeux autour d’une élaboration de chartes d’entreprise sur l’IA puis sur ce qu’implique l’utilisation de celles-ci.

+	-
Sur l’écriture de la charte	
Formalisation de l’engagement	Ambiguïté, interprétation

¹¹ <https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/>

Influence des textes législatifs	Simple rappel de la loi
Principes largement représentés	Principes sous représentés
Sur l'utilité de la charte	
<i>Dans les rapports internes</i>	
Ligne de conduite Renforcement des valeurs de l'entreprise Avantage concurrentiel pour les salariés	Caractère unilatéral, non opposabilité et absence de sanction (droit souple)
<i>Dans les rapports externes</i>	
Démarche proactive, volonté d'autorégulation Recherche en amont de la législation Avantage concurrentiel pour les clients Réduction de la méfiance envers les nouvelles technologies	Caractère unilatéral, non opposabilité et absence de sanction (droit souple) Simple vitrine Outil de communication

Figure 2 : Tableau récapitulatif de l'analyse de la carte heuristique "Chartes d'entreprises sur l'Intelligence Artificielle"

Tout d'abord, l'intelligence artificielle n'est que le reflet des données servant à son élaboration (*garbage in, garbage out*). Les chartes d'entreprises qui encadrent la vie en entreprise de l'intelligence artificielle en font tout autant. L'écriture de la charte témoigne de nombreuses réalités. Cette étape formalise le questionnement entrepris par les sphères décisionnelles d'une entreprise tout en s'inspirant des textes en vigueur. Toutefois, ce qui se laisse paraître comme engagement fort d'une direction ne va parfois pas plus loin qu'un simple rappel de la loi. Prenons l'exemple de la protection des données. De nombreuses contraintes existent notamment via le célèbre Règlement Général sur la Protection des

Données. Certaines chartes transforment une exigence légale en outil de communication à propos d'un soi-disant engagement. Toute personne - morale ou physique - est cependant déjà tenue de respecter la loi.

La législation influence tout naturellement la répartition des engagements des chartes par principe. Certains sont plus ou moins représentés. Le respect de la vie privée l'est largement, nous venons de l'évoquer. Quelques entreprises vont toutefois plus loin. Nous nous souviendrons de l'engagement clair derrière un bien-être social. Malgré tout, les chartes d'entreprises défendent peu certains principes comme la responsabilité. Doit-on y voir un lien avec le manque de législation ou est-ce plus largement l'ambiguïté de certains termes autorisant une très large interprétation ? Les deux ne sont pas incompatibles.

Sur l'utilité des chartes maintenant. Nous nous penchons dans un premier temps sur les rapports internes à l'entreprise, influencés par la rédaction d'une charte. Comme souligné précédemment, la charte relève du droit souple. Même dans sa version la plus lourde, il s'agit d'un contrat. Dans les autres cas, aucune sanction ne peut être prise si la charte est outrepassée. La charte ne représente qu'un engagement de façade. Il va sans dire que le manager peut s'appuyer sur une charte pour guider son action telle une source d'inspiration, mais elle ne peut lui servir de justification à la sanction. La charte permet une nouvelle fois de formaliser une pensée, de témoigner d'un engagement. Décider de rentrer dans ce processus d'écriture, c'est aussi chercher à réunir des collaborateurs autour de valeurs communes. Gardons à l'esprit que les métiers de l'IA sont en plein devenir. Les talents se font rares. Afficher une charte clairement établie crée une fenêtre sur l'extérieur. Un talent qui se pose la question d'intégrer l'une ou l'autre entreprise cherche un critère différenciant entre deux structures et essaye potentiellement de connaître les valeurs défendues par ces structures. La charte d'entreprise accorde un nivellement concurrentiel entre plusieurs firmes, précieux dans les métiers de l'IA.

C'est sur ce rôle de fenêtre sur l'extérieur que nous allons désormais nous attarder. L'IA est un sujet de méfiance pour la population. Qu'une entreprise manifeste d'un engagement et de principes clairs atteste des efforts de pédagogie que produit la structure. La société entre dans une démarche proactive et souhaite uniformiser le travail de ses équipes. Il faut toutefois rester vigilant. Afficher de bonnes volontés ne signifie pas que ces dernières sont appliquées de manière effective. La charte d'entreprise ne doit pas rester vernis de surface ou outil communicationnel. Certes, le pouvoir du client se fait de plus en plus sentir.

À service rendu égal, le client peut choisir un fournisseur en fonction des valeurs défendues. Tout comme le futur talent, un effort de transparence permet de toucher une nouvelle clientèle ou de préserver les clients actuels. La rédaction et l'adoption d'une charte d'entreprise sur l'utilisation de l'IA va bien plus loin que la rédaction d'un traité de bonne conduite, mais témoigne d'un questionnement de la sphère stratégique qui se traduit nécessairement sur la gestion d'une entreprise.

La charte d'entreprise se retrouve aux confins de l'application de la législation en vigueur, du support de communication (interne et externe) et de la *soft law*. L'outil en lui-même s'avère toutefois insuffisant. Il ressort de cette étude que la charte d'entreprise se montre comme une démarche intéressante, mais qui manque de poids. La proposition existe mais ne dispose pas d'instruments opposables pour vérifier son efficacité. Pour que l'initiative se transforme en réel outil stratégique, il est nécessaire de créer des outils de mesure et qu'une gouvernance réelle de l'Intelligence Artificielle soit mise en place dans l'entreprise. Les questions éthiques autour de l'Intelligence Artificielle étant un enjeu d'avenir, il semble également important de permettre une mesure de son utilisation par le grand public (ainsi qu'un moyen de comparaison entre les entreprises) par la création de labels ou de certifications.

Pour conclure, la rédaction d'une charte témoigne d'un questionnement de la sphère stratégique qui se traduit nécessairement sur la gestion d'une entreprise. L'outil en lui-même est toutefois insuffisant. Nous sommes donc favorables à l'utilisation de chartes sur l'IA au sein des entreprises, à condition que soient introduits des outils de mesure pour l'évaluation du respect de ces chartes et afin d'en renforcer la transparence et que soient traduits les principes éthiques dans la gouvernance. Nous proposons également que soient valorisées les bonnes pratiques à travers la création de labels et certifications, comme on en trouve pour l'environnement.

Définition des principes des institutions retenues :

// AI HLEG = High Level Expert Group on AI //

juin 2018 - avril 2019

Ethics Guidelines for Trustworthy AI

Ce sont les 7 principes éthiques **européens** concernant l'IA.

Action et contrôle humain : les systèmes d'IA devraient être les vecteurs de sociétés équitables en se mettant au service de l'humain et des droits fondamentaux, sans restreindre ou dévoyer l'autonomie humaine.

Robustesse technique et sécurité : une IA digne de confiance nécessite des algorithmes suffisamment sûrs et fiables pour gérer les erreurs ou les incohérences dans toutes les phases du cycle de vie des systèmes d'IA.

Respect de la vie privée et gouvernance des données : il faut que les citoyens aient la maîtrise totale de leurs données personnelles et que les données les concernant ne soient pas utilisées contre eux à des fins préjudiciables ou discriminatoires.

Transparence : la traçabilité des systèmes d'IA doit être assurée

Diversité, non-discrimination et équité : les systèmes d'IA devraient prendre en compte tout l'éventail des capacités, aptitudes et besoins humains, et leur accessibilité devraient être garantie.

Bien-être social et environnemental : les systèmes d'IA devraient être utilisés pour soutenir des évolutions sociales positives et renforcer la durabilité et la responsabilité écologique.

Responsabilité : il convient de mettre en place des mécanismes pour garantir la responsabilité à l'égard des systèmes d'IA et de leurs résultats, et de les soumettre à une obligation de rendre des comptes.

// Impact AI //

février 2020 - décembre 2020

IA digne de confiance - Construire une gouvernance adaptée à chaque entreprise

Collectif de réflexion et d'action constitué de grandes entreprises, d'entreprises de services numériques, de sociétés de conseil en stratégie, de startups et d'écoles.

Dignité : les systèmes d'intelligence artificielle devraient être les vecteurs de sociétés équitables en se mettant au service de l'humain et des droits fondamentaux, sans restreindre ou dévoyer l'autonomie des individus.

Robustesse : une IA digne de confiance nécessite des algorithmes suffisamment sûrs et fiables pour gérer les erreurs ou les incohérences à toutes les étapes du cycle de vie d'un outil d'intelligence artificielle.

Gouvernance des données : ensemble des mesures qui visent à s'assurer de la qualité et de l'intégrité des données utilisées ainsi que de leur pertinence dans le domaine d'application considéré. Il s'agit aussi d'établir des protocoles d'accès à ces données et des règles pour les traiter qui garantissent le respect de la vie privée.

Transparence : les algorithmes d'apprentissage sont dits transparents lorsque le fonctionnement interne des modèles d'apprentissage est rendu public ; ils répondent à l'exigence d'explicabilité quand ces mêmes éléments restent intelligibles, c'est-à-dire qu'il est possible de comprendre à quelles règles ils obéissent.

Équité : alors que l'on attend des résultats d'un algorithme de machine learning une neutralité parfaite, il arrive que ceux-ci soient biaisés. Un biais est un préjugé en faveur ou en défaveur d'une chose, d'une personne ou d'un groupe par rapport à un autre et qui est considéré comme injuste. L'IA équitable, débarrassée des biais, peut contribuer à réduire les inégalités et à bâtir une société plus inclusive.

Développement durable et bien-être : les systèmes d'intelligence artificielle devraient être utilisés pour soutenir les évolutions sociales positives et renforcer la durabilité et la responsabilité écologique.

Responsabilité : le développement de l'intelligence artificielle doit s'accompagner de garanties suffisantes pour minimiser le risque de dommages qu'elle peut causer. Il convient de mettre en place des mécanismes à cette fin et de les soumettre à l'obligation de rendre des comptes.

// Partnership on AI //

Empower people : We will seek to ensure that AI technologies benefit and empower as many people as possible.

Engage : We will educate and listen to the public and actively engage stakeholders to seek their feedback on our focus, inform them of our work, and address their questions.

Research and dialogue : We are committed to open research and dialogue on the ethical, social, economic, and legal implications of AI. We believe that AI research and development efforts need to be actively engaged with and accountable to a broad range of stakeholders.

Address concerns : We will engage with and have representation from stakeholders in the business community to help ensure that domain-specific concerns and opportunities are understood and addressed.

Understandability : We believe that it is important for the operation of AI systems to be understandable and interpretable by people, for purposes of explaining the technology.

Cooperation and trust : We strive to create a culture of cooperation, trust, and openness

among AI scientists and engineers to help us all better achieve these goals.

We will work to maximize the benefits and address the potential challenges of AI technologies, by:

- Working to protect the privacy and security of individuals.
- Striving to understand and respect the interests of all parties that may be impacted by AI advances.
- Working to ensure that AI research and engineering communities remain socially responsible, sensitive, and engaged directly with the potential influences of AI technologies on wider society.
- Ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.
- Opposing development and use of AI technologies that would violate international conventions or human rights, and promoting safeguards and technologies that do no harm.

Différentes chartes d'entreprise :

// ARBORUS //

Mixité et diversité : promouvoir la mixité et la diversité dans les équipes qui travaillent sur des solutions à base d'IA.

Correction des biais : s'organiser pour évaluer et réagir à toutes formes de discrimination qui pourraient résulter de données biaisées ou stéréotypées.

Veiller à la qualité des données utilisées pour garantir des systèmes les plus équitables possible : une donnée unifiée, cohérente, vérifiée, traçable et exploitable.

Formation des concepteurs d'IA : former pour sensibiliser et responsabiliser les concepteurs, développeurs et tous les acteurs impliqués dans la fabrication de l'IA, aux stéréotypes, aux biais pouvant générer des discriminations.

Sensibilisation aux biais : sensibiliser les prescripteurs des solutions à base d'IA (RH, finances, relations clients, marketing ...), aux risques de biais et stéréotypes pouvant générer des discriminations et intégrer dans les cahiers des charges des points de contrôle et d'évaluation itérative.

Sélection des fournisseurs : veiller à bien choisir les fournisseurs et les évaluer de manière itérative afin de s'assurer que toute la chaîne de valeur de l'IA soit non discriminatoire.

Contrôle et adaptation des processus : contrôler les solutions à base d'IA et adapter en continu les processus.

// Crédit Agricole //

Le groupe Crédit Agricole, première banque des Français, a souhaité se différencier et prendre une position forte pour protéger les données personnelles de ses clients. Leur utilisation encadrée se fera en toute sécurité au sein du Groupe et avec pour seul objectif l'amélioration du service apporté aux clients. Cette Charte co-construite avec les clients du Crédit Agricole repose sur cinq principes :

Utilité et loyauté : le Groupe s'engage à utiliser les données dans l'intérêt de ses clients pour proposer un conseil et des offres personnalisées, un service de plus grande qualité et fournir tous les éléments pour les aider à prendre les meilleures décisions.

Éthique : le Groupe s'engage à agir avec éthique et responsabilité en matière de données personnelles et s'interdit de vendre les données personnelles de ses clients.

Transparence et pédagogie : le Groupe s'engage à expliquer de manière claire et

transparente ce qui est fait en termes d'utilisation des données.

Maîtrise aux mains des clients : le Groupe s'engage à laisser la maîtrise de leurs données à ses clients et l'utilisation qui en est faite. Il annonce le lancement d'un projet de création en 2017 d'un espace client sécurisé dédié à la gestion de ses données.

Sécurité des données : la sécurité des données reste la priorité du Groupe.

« Dans un contexte où chacun s'interroge sur la protection de sa vie privée et de ses données personnelles, notre Charte s'appuie sur une conviction forte : l'utilisation des données, dans un cadre défini, n'est légitime que si elle sert l'intérêt de nos clients. Nous souhaitons nous inscrire comme le « vrai partenaire » de nos clients et maintenir le capital confiance élevé du Groupe. » résume Bertrand Corbeau, Directeur général adjoint de Crédit Agricole S.A. en charge du Pôle Développement, Client et Innovation.

Cette Charte sera diffusée à l'ensemble des collaborateurs et remise aux clients du Groupe par leur conseiller dans les moments clés de la relation.

// MAIF //

Charte MAIF POUR UN MONDE NUMÉRIQUE RÉSOLUMENT HUMAIN ET ÉTHIQUE

Enrichir : Le progrès technique doit rester au service de l'Homme. Nous utilisons les technologies, en particulier l'intelligence artificielle, au service de l'Homme, afin d'enrichir aussi bien la qualité de la relation avec nos sociétaires que les métiers et savoir-faire de nos équipes.

Piloter : L'Homme doit toujours conserver la maîtrise de la technique. Nous garantissons que les technologies et les algorithmes que nous utilisons sont placés sous le contrôle de l'Homme.

Mettre en capacité : Pour que le numérique soit au service de la confiance. Nous promouvons une nouvelle politique de la donnée visant à fournir aux individus la connaissance, le contrôle effectif et la maîtrise de l'usage des données qui les concernent, afin de développer leur pouvoir d'agir.

// MICROSOFT AI PRINCIPLES //

Nous vous invitons à explorer les vidéos sur le site afin d'approfondir les principes énoncés.

Fairness : AI systems should treat all people fairly

Reliability & Safety : AI systems should perform reliably and safely

Privacy & Security : AI systems should be secure and respect privacy

Inclusiveness : AI systems should empower everyone and engage people

Transparency : AI systems should be understandable

Accountability : People should be accountable for AI systems

// SAP //

septembre 2018

SAP's Guiding Principles for Artificial Intelligence

SAP est la première entreprise tech européenne à créer un panel consultatif d'éthique pour l'intelligence artificielle.

We are driven by our values : We recognize that, like with any technology, there is scope for AI to be used in ways that are not aligned with these guiding principles and the operational guidelines we are developing. In developing AI software we will remain **true to our Human Rights Commitment Statement, the UN Guiding Principles on Business**

and Human Rights, laws, and widely accepted international norms. Wherever necessary, our **AI Ethics Steering Committee** will serve to advise our teams on how specific use-cases are affected by these guiding principles. Where there is a conflict with our principles, we will endeavor to prevent the inappropriate use of our technology.

We design for people : We strive to create AI software systems that are inclusive and that seek to empower and augment the talents of our diverse usership. By providing **human-centered user experiences** through augmentative and intuitive technologies, we leverage AI to support people in maximizing their potential. To achieve this, we **design our systems closely with users in a collaborative, multidisciplinary, and demographically diverse environment.**

We enable businesses beyond bias : Bias can negatively impact AI software and, in turn, individuals and our customers. This is particularly the case when there is a **risk of causing discrimination or of unjustly impacting underrepresented groups.** We, therefore, require our technical teams to gain a deep understanding of the business problems they are trying to solve, and the data quality this demands. We seek to **increase the diversity and interdisciplinarity of our teams**, and we are investigating new technical methods for mitigating biases. We are also deeply committed to supporting our customers in building even more diverse businesses by leveraging AI to build products that help move business beyond bias.

We strive for transparency and integrity in all that we do : **Our systems are held to specific standards** in accordance with their level of technical ability and intended usage. Their **input, capabilities, intended purpose, and limitations** will be communicated clearly to our customers, and we provide means for oversight and control by customers and users. They are, and will always remain, in control of the deployment of our products. We actively support industry collaboration and will conduct research to further system transparency.

We uphold quality and safety standards : As with any of our products, our AI software is subject to our **quality assurance process**, which we continuously adapt when necessary. Our AI software undergoes thorough testing under real-world scenarios to firmly validate

they are fit for purpose and that the product specifications are met. We work closely with our customers and users to uphold and further improve our systems' quality, safety, reliability, and security.

We place data protection and privacy at our core : **Data protection and privacy are a corporate requirement** and at the core of every product and service. We communicate clearly how, why, where, and when customer and anonymized user data is used in our AI software.

We engage with the wider societal challenges of AI : While we have control, to a large extent, over the preceding areas, there are numerous emerging challenges that require a much broader discourse across industries, disciplines, borders, and cultural, philosophical, and religious traditions. These include, but are not limited to, questions concerning:

Economic impact, such as how industry and society can collaborate to prepare students and workers for an AI economy and how society may need to adapt means of economic redistribution, social safety, and economic development.

Social impact, such as the value and meaning of work for people and the potential role of AI software as social companions and caretakers.

Normative questions around how AI should confront ethical dilemmas and what applications of AI, specifically with regards to security and safety, should be considered permissible.

[// Google //](#)

AI at Google: our principles

Be socially beneficial : The expanded reach of new technologies increasingly touches society as a whole. Advances in AI will have transformative impacts in a wide range of fields, including healthcare, security, energy, transportation, manufacturing, and entertainment. As we consider potential development and uses of AI technologies, we will

take into account a broad range of social and economic factors, and will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides.

Avoid creating or reinforcing unfair bias : AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. We recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.

Be built and tested for safety : We will continue to develop and apply strong safety and security practices to avoid unintended results that create risks of harm. We will design our AI systems to be appropriately cautious, and seek to develop them in accordance with best practices in AI safety research. In appropriate cases, we will test AI technologies in constrained environments and monitor their operation after deployment.

Be accountable to people : We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control.

Incorporate privacy design principles : We will incorporate our privacy principles in the development and use of our AI technologies. We will give opportunity for notice and consent, encourage architectures with privacy safeguards, and provide appropriate transparency and control over the use of data.

Uphold high standards of scientific excellence : Technological innovation is rooted in the scientific method and a commitment to open inquiry, intellectual rigor, integrity, and collaboration. AI tools have the potential to unlock new realms of scientific research and knowledge in critical domains like biology, chemistry, medicine, and environmental sciences. We aspire to high standards of scientific excellence as we work to progress AI development. We will work with a range of stakeholders to promote thoughtful leadership in this area, drawing on scientifically rigorous and multidisciplinary approaches. And we

will responsibly share AI knowledge by publishing educational materials, best practices, and research that enable more people to develop useful AI applications.

Be made available for uses that accord with these principles : Many technologies have multiple uses. We will work to limit potentially harmful or abusive applications. As we develop and deploy AI technologies, we will evaluate likely uses in light of the following factors:

Primary purpose and use: the primary purpose and likely use of a technology and application, including how closely the solution is related to or adaptable to a harmful use

Nature and uniqueness: whether we are making available technology that is unique or more generally available

Scale: whether the use of this technology will have significant impact

Nature of Google's involvement: whether we are providing general-purpose tools, integrating tools for customers, or developing custom solutions

[// SONY //](#)

Supporting Creative Life Styles and Building a Better Society : Through advancing its AI-related R&D and promoting the utilization of AI in a manner harmonized with society, Sony aims to support the exploration of the potential for each individual to empower their lives, and to contribute to enrichment of our culture and push our civilization forward by providing novel and creative types of kando. Sony will engage in sustainable social development and endeavor to utilize the power of AI for contributing to global problem-solving and for the development of a peaceful and sustainable society.

Stakeholder Engagement : In order to solve the challenges arising from use of AI while striving for better AI utilization, Sony will seriously consider the interests and concerns of various stakeholders including its customers and creators, and proactively advance a dialogue with related industries, organizations, academic communities and more. For this purpose, Sony will construct the appropriate channels for ensuring that the content and results of these discussions are provided to officers and employees, including researchers

and developers, who are involved in the corresponding businesses, as well as for ensuring further engagement with its various stakeholders.

Provision of Trusted Products and Services : Sony understands the need for safety when dealing with products and services utilizing AI and will continue to respond to security risks such as unauthorized access. AI systems may utilize statistical or probabilistic methods to achieve results. In the interest of Sony's customers and to maintain their trust, Sony will design whole systems with an awareness of the responsibility associated with the characteristics of such methods.

Privacy Protection : Sony, in compliance with laws and regulations as well as applicable internal rules and policies, seeks to enhance the security and protection of customers' personal data acquired via products and services utilizing AI, and build an environment where said personal data is processed in ways that respect the intention and trust of customers.

Respect for Fairness : In its utilization of AI, Sony will respect diversity and human rights of its customers and other stakeholders without any discrimination while striving to contribute to the resolution of social problems through its activities in its own and related industries.

Pursuit of Transparency : During the planning and design stages for its products and services that utilize AI, Sony will strive to introduce methods of capturing the reasoning behind the decisions made by AI utilized in said products and services. Additionally, it will endeavor to provide intelligible explanations and information to customers about the possible impact of using these products and services.

The Evolution of AI and Ongoing Education : People's lives have continuously changed with the advance in technology across history. Sony will be cognizant of the effects and impact of products and services that utilize AI on society and will proactively work to contribute to developing AI to create a better society and foster human talent capable of shaping our collective bright future through R&D and/or utilization of AI.